

Factors of choice in urban mobility: discrete choice models for the Asprela university campus

Luís Miguel Ferreira Meireles

Master's Dissertation

Supervisor: Prof. Beatriz Brito Oliveira



Mestrado em Engenharia e Gestão Industrial

2025-01-26

Abstract

This study explores urban mobility within the Asprela university campus in Porto as part of the "Modeling shared mobility: advanced demand modeling and learning for the sustainable shared mobility of the future" (MOSH) project, funded by FCT (Fundação para a Ciência e Tecnologia) under grant 2022.03138.PTDC. By focusing on the decision-making processes influencing mode choice, the research employs discrete choice models (DCM) to examine the most determinant factors that individuals commuting in this region consider when choosing how to travel. Three groups condense these factors: socio-demographic, objective, and subjective factors. The research was initiated by analyzing data from a Revealed Preferences (RP) survey conducted within the scope of the MOSH project. This survey covered a population characterized by many younger individuals, mostly university students and workers. Subsequently, various modeling attempts using distinct types of DCMs were performed to estimate and interpret the impact of the specified factors. In descending order of popularity, choice was considered one of the following alternatives: bus, metro, walking, own car, riding with someone, ride-hailing platform, bicycle/scooter, and taxi. The findings intend to inform stakeholders in urban planning and mobility. Even though this study does not aim to propose any specific policy, it can hopefully shed new light on and influence future transportation policies and sustainable commuting practices in metropolitan areas, university campuses, or hospital complexes. The proposed analysis unveiled interesting insights, among others that were less surprising. For instance, the results demonstrate that individuals systematically prefer modes like own car and metro over buses. Owning a subscription discourages walking trips, while higher income levels make public transportation less attractive. Furthermore, individuals favor metro over bus for highly regular trips. Individuals are more willing to pay to travel with their car or resort to ride-hailing. Marginal increases in waiting/walking times impact more than similar increases in trip duration. Regarding individual perception, lower satisfaction is expected from public transportation; individuals also substantially avoid ride-hailing. A counterintuitive finding is that walking seems more likely to be chosen for longer trips as long as it is feasible. Besides, the investigation identified several methodological and practical implications. While the multinomial logit model (MNL) still provides one of the most reliable baseline options to understand underlying choice patterns, slightly more sophisticated approaches like the latent class model (LCM) offer a deeper exploration of more complex choice dynamics by partially relaxing the independence of irrelevant alternatives (IIA) property and introducing heterogeneity to model individual preferences. Nested extensions of logit (NL) bridge similar gaps, yet they did not yield significant results due to computational issues. Finally, the results emphasize the importance of employing robust validation techniques in future research on mode choice.

Resumo

Este estudo explora a mobilidade urbana no contexto do pólo universitário da Asprela, situado no Porto, enquanto parte do projeto "Modeling shared mobility: advanced demand modeling and learning for the sustainable shared mobility of the future" (MOSH), financiado pela FCT (Fundação para a Ciência e Tecnologia) sob a bolsa 2022.03138.PTDC. Ao focar-se no processos de decisão por trás da escolha modal, a investigação utiliza modelos de escolha discreta (MED) para examinar os fatores mais determinantes que os indivíduos que pendulam nesta região consideram aquando da escolha de como viajar. Três grupos condensam estes fatores: fatores socio-demográficos, objetivos, e subjetivos. A investigação arrancou com a análise dos dados obtidos a partir de um inquérito de preferências reveladas realizado no âmbito do projeto MOSH. Este inquérito abrangeu uma população caracterizada por bastantes indivíduos mais jovens, maioritariamente estudantes universitários e trabalhadores. Subsequentemente, várias tentativas de modelagem utilizando tipos distintos de MEDs foram realizadas para estimar e interpretar o impacto dos fatores especificados. Em ordem decrescente de popularidade, a escolha foi considerada uma das seguintes alternativas: autocarro, metro, caminhar, carro próprio, boleia, plataforma de transporte por aplicativo, bicicleta/scooter, e táxi. As presentes conclusões visam informar intervenientes no planeamento e mobilidade urbanos. Embora este estudo não tenha como objetivo propôr qualquer política específica, espera-se que possa oferecer uma nova perspetiva e influenciar futuras políticas de transporte e práticas de mobilidade sustentáveis em áreas metropolitanas, pólos universitários, ou complexos hospitalares. A análise proposta revelou conclusões de interesse, entre outras menos surpreendentes. Por exemplo, os resultados demonstram que os indivíduos preferem sistematicamente modos como carro próprio e metro relativamente ao autocarro. Possuir uma subscrição desencoraja viagens a pé, enquanto níveis de rendimento superiores fazem do transporte público menos atrativo. Ademais, os indivíduos preferem o metro ao autocarro para viagens mais frequentes. Indivíduos estão mais dispostos a pagar para viajar com o seu carro ou recorrer a uma plataforma de transporte por aplicativo. Incrementos marginais nos tempos de espera/caminhada impactam mais do que incrementos similares na duração da viagem. Em relação à perceção individual, é esperada satisfação inferior ao usar transportes públicos, enquanto plataformas de transporte por aplicativo são substancialmente evitadas. Uma conclusão contra-intuitiva é que caminhar aparenta ser mais provável para viagens mais longas, desde que exequível. Além disso, a investigação identificou várias implicações metodológicas e práticas. Embora o modelo logit multinomial continue a ser uma das opções de referência mais fiáveis para compreender os padrões de escolha subjacentes, abordagens ligeiramente mais sofisticadas como o modelo de classes latentes oferecem uma exploração mais profunda das dinâmicas de escolha mais complexas ao relaxar parcialmente a propriedade de independência das alternativas irrelevantes e introduzir heterogeneidade para modelar preferências individuais. As extensões aninhadas do logit preenchem lacunas semelhantes, no entanto, não produziram resultados significativos devido a problemas computacionais. Por fim, os resultados enfatizam a importância de recorrer a técnicas de validação robustas em investigações de escolha modal futuras.

Acknowledgments

I sincerely thank my parents for the unwavering effort and patience invested in me. To the woman who accompanied me every single day throughout this adventure, I am deeply grateful. To my brother, who still has the pleasure of living with me after 25 years, thank you for your companionship. To all the friends who carried me along the way, your encouragement has meant the world to me. Finally, my most profound appreciation goes to Professor Beatriz Brito Oliveira for all the invaluable motivation and guidance, week after week, bringing out the best in me. Thank you for your constant availability, rigor, and shared knowledge. I am profoundly grateful for the opportunity you provided me, and I extend my heartfelt congratulations for your outstanding work on the MOSH project.

"You just stay alert, do you work, don't worry about the world going by. It doesn't mean that you can sit around and not do anything. You've got to get your butt in gear and do it. And don't take no for an answer. Translate those ideas to cinema, or to a painting, or to whatever. Figure out a way to get it done."

David Lynch

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Project description	2
1.3	Project objectives	2
1.4	Dissertation structure	3
2	Literature review	5
2.1	Transportation mode choice	5
2.1.1	Subjective factors	6
2.1.2	Objective factors	8
2.1.3	Socio-demographic, economic, and cultural factors	11
2.1.4	Built environment	14
2.2	Discrete choice models	16
2.2.1	Multinomial logit model	17
2.2.2	Mixed logit model	18
2.2.3	Latent class model	20
2.2.4	Nested logit model	21
2.2.5	Cross-nested logit model	22
3	Methodology and data	25
3.1	Survey overview	26
3.1.1	Socio-demographics	31
3.1.2	Transportation mode choice	32
3.1.3	Objective outcomes of the trip	34
3.1.4	Experience outcomes of the trip	37
3.2	Imputation method and validation heuristic	40
3.2.1	Validation rationale	43
3.2.2	Validation metrics	45
3.2.3	Validation results	49
3.3	Multinomial logit model	51
3.3.1	Definition of the sets of factors	56
3.3.2	Validation results	58
3.4	Nested and cross-nested logit models	61
3.5	Latent class model	64
3.5.1	Validation results	65

4	Results and discussion	69
4.1	Validation results: multinomial logit vs. latent class	69
4.2	Estimation results and analysis	71
4.2.1	Alternative-specific constants	72
4.2.2	Socio-demographic factors	75
4.2.3	Objective factors	78
4.2.4	Subjective factors	85
4.2.5	Willingness to pay	87
5	Conclusions and future research	91
A	Descriptive statistics	101
B	Parameter estimation results	113

Acronyms and Symbols

ACC	Accuracy (same as True Positive Rate, in the present context)
AIC	Akaike Information Criterion
APE	Absolute Percentage Error
ASC	Alternative Specific Constant
BIC	Bayesian Information Criterion
BS1	Backward Selection step 1 set
BS2	Backward Selection step 2 set
CNL	Cross-Nested Logit Model
CV	Cross-Validation
DCM	Discrete Choice Model
EV	Extreme Value Distribution
F1	F1-score
GEV	Generalized Extreme Value
HD	Hypotheses-Driven set
HOV	Holdout Validation
IIA	Independence of Irrelevant Alternatives
IID	Independent and Identically Distributed
ISV	In-Sample Validation
LCM	Latent Class Model
LCM1	Class 1 of Latent Class Model
LCM2	Class 2 of Latent Class Model
LL	Log-Likelihood
$LL_e(\hat{\beta}^e)$	Log-Likelihood of the model estimated on the entire sample
$LL_e(0)$	Null Log-Likelihood associated with the entire sample
$LL_t(\hat{\beta}^t)$	Log-Likelihood of the model estimated on the training sample
$LL_t(0)$	Null Log-Likelihood associated with the training sample
$LL_v(\hat{\beta}^t)$	Log-Likelihood of the model estimated on the training sample applied to the validation sample
$LL_v(0)$	Null Log-Likelihood associated with the validation sample
LRT	Likelihood Ratio Test
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MAR	Missing At Random
MCAR	Missing Completely At Random
MEI	Median Imputation
MLE	Maximum Likelihood Estimation
MICE	Multiple Imputation by Chained Equations
MENA	Middle East and North Africa

MNAR	Missing Not At Random
MNL	Multinomial Logit Model
MMNL	Mixed Logit Model
MOI	Mode Imputation
MOSH	"Modeling shared mobility: advanced demand modeling and learning for the sustainable shared mobility of the future"
MSLE	Maximum Simulated Likelihood
MSM	Method of Simulated Moments
NL	Nested Logit Model
NLL	Null Log-likelihood
OSV	Out-of-Sample Validation
PREC	Precision
RF	Random Forest
RFI	Random Forest Imputation
RP	Revealed Preferences
RUM	Random Utility Model
SEM	Structural Equation Model(ing)
SP	Stated Preferences
TPR	True Positive Rate or Recall (same as Accuracy, in the present context)
TT1	Threshold Test step 1 set
TT2	Threshold Test step 2 set
VT	Vuong Test
VTTS	Value of Travel Time Savings
WTP	Willingness To Pay
ρ^2	Likelihood Ratio Index or McFadden's R^2
$\bar{\rho}^2$	Adjusted Likelihood Ratio Index or Adjusted McFadden's R^2
%CR	Percentage of Clearly Right Predictions
%CW	Percentage of Clearly Wrong Predictions
%U	Percentage of Unclear Predictions

List of Figures

3.1	Methodology and discussion outline.	26
3.2	Distribution of mode availability vs. mode choice.	34
3.3	Distribution of the 20 most frequent mode availability combinations vs. mode choice.	35
3.4	Median of the answers to 5-point Likert scale questions regarding experience, by mode choice.	38
3.5	Mean of the answers to 10-point Likert scale questions regarding experience, by mode choice.	39
A.1	Univariate distribution of participant's age.	102
A.2	Distribution of average participant's age, by mode choice.	102
A.3	Univariate distribution of participant's sex.	103
A.4	Univariate distribution of participant's education level.	103
A.5	Univariate distribution of participant's household composition.	104
A.6	Distribution of participant's subscription owned, by mode choice.	104
A.7	Distribution of participant's monthly income level, by mode choice.	105
A.8	Distribution of whether the participant's trip was regular, by mode choice.	105
A.9	Distribution of participant's trip regularity, by mode choice.	106
A.10	Univariate distribution of participant's trip purpose.	106
A.11	Distribution of participant's trip purpose, by mode choice.	107
A.12	Distribution of average number of modes used, by mode choice.	107
A.13	Distribution of number of modes used, by mode choice.	108
A.14	Distribution of trip duration, by mode choice.	108
A.15	Distribution of average trip duration vs. total journey duration, by mode choice.	109
A.16	Distribution of trip cost, by mode choice.	109
A.17	Distribution of walking time, by mode choice.	110
A.18	Distribution of waiting time, by mode choice.	110
A.19	Pattern of missing data.	111

List of Tables

3.1	Breakdown of MOSH project's survey. Only revealed preferences sections were considered.	27
3.2	Results of Pearson's chi-squared test of independence using choice as the target variable.	28
3.3	Results of Pearson's correlation coefficients for non-choice pairs.	29
3.4	Results of Pearson's correlation coefficients for choice-only pairs.	30
3.5	Factors included in the modeling phase.	43
3.6	In-sample validation results from the datasets generated with each imputation technique.	49
3.7	Out-of-sample validation results from the datasets generated with each imputation technique.	50
3.8	Validation sample's market shares (%) from the datasets generated with each imputation technique.	50
3.9	Sets of factors considered for parameter estimation.	57
3.10	In-sample validation results from MNL using all sets of factors, except HD. . . .	59
3.11	Out-of-sample validation results from MNL using all sets of factors, except HD. .	60
3.12	Validation sample's market shares (%) from MNL using all sets of factors, except HD.	60
3.13	In-sample validation results from LCM using sets of factors BS2 and TS2.	65
3.14	Out-of-sample validation results from LCM using sets of factors BS2 and TT2. .	66
3.15	Validation sample's market shares (%) from LCM using sets of factors BS2 and TT2.	66
4.1	In-sample validation results from LCM and MNL using set of factors TT2.	69
4.2	Out-of-sample validation results from LCM and MNL using set of factors TT2. .	70
4.3	Validation sample's market shares (%) from LCM and MNL using set of factors TT2.	71
4.4	Parameter estimation results from MNL using set of factors "Threshold test step 2".	73
4.5	Willingness to pay in euros (€) for the parameter estimation results from MNL using set of factors TT2.	89
B.1	Parameter estimation results from LCM - Class 1 using set of factors "Threshold test step 2". The estimated value of H_{n1} is 0.397. Coefficients are scaled by a factor of 0.1 for compact representation.	114
B.2	Parameter estimation results from LCM - Class 2 using set of factors "Threshold test step 2". By inference, the value of H_{n2} is 0.603. Coefficients are scaled by a factor of 0.1 for compact representation.	115

Chapter 1

Introduction

This chapter introduces the research project, setting the foundation for the present Master's dissertation. It starts by characterizing the context and motivation for the study, followed by a slightly more detailed description of the project itself, and concludes with its outline and proposed structure of the document.

1.1 Motivation

The theme of the present dissertation is mode choice analysis, with a special emphasis on urban settings. Urban mobility typically represents a critical issue in modern urban planning, with direct implications for sustainability and the quality of life of city inhabitants. The document explores the most essential aspects of urban mobility, particularly their significance and impact within the context of Asprela, a key university and hospital hub in Porto.

The relevance of this research stems from the will to improve the understanding of human behavior regarding the decision-making process inherent to choosing which mode to travel. Consequently, the study focuses mainly on the factors most expected to influence these choices. While its focus is primarily on other areas, it still offers valuable insights into urban planning and sustainable transportation policies. These findings expect to motivate other researchers and policy-makers to evaluate their practical consequences further and pursue the implementation of different methods and novel ideas in examining similar cities and/or countries. The interest of such findings to raise awareness of the average individual is also noteworthy.

This project was conducted in collaboration with the Institute for Systems and Computer Engineering, Technology and Science (INESC TEC) as a segment of a larger project hosted by the institution, titled "Modeling shared mobility: advanced demand modeling and learning for the sustainable shared mobility of the future" (MOSH) and funded by FCT (Fundação para a Ciência e Tecnologia) under grant 2022.03138.PTDC.

The context underscores the need for a basic understanding of commuting behaviors typical of metropolitan areas, namely academic or hospital environments, in addition to an entry-level knowledge of discrete choice analysis.

1.2 Project description

The Asprela university campus, characterized by a high commuting demand and a diverse population, including a large percentage of students and workers, provides a rich, unique, and almost ideal environment to study urban traveling patterns and their underlying dynamics. The project's scope extends to understanding the interplay between individual preferences and social constraints. The goal is to describe hypotheses about the decision-making process behind mode choice, which can benefit urban planning and mobility stakeholders.

Notwithstanding the importance of built environment-related factors for mode choice in the present context, this project focuses solely on investigating what is referred to as objective, subjective, and socio-demographic factors. By doing so, the study employed discrete choice modeling methodologies to identify, quantify, and interpret them. Discrete Choice Models (DCMs) are used in mode choice analysis to predict the likelihood of choosing between different traveling alternatives and interpret individual preferences. They are based on random utility theory, i.e., assume that individuals choose the alternative yielding the highest personal satisfaction, also known as utility. Common types of DCMs include the Multinomial Logit Model (MNL) and nested extensions, such as the Nested Logit Model (NL). Some, like the Mixed Logit Model (MMNL), can account for population heterogeneity. Nonetheless, these models are often subject to limitations, such as biases found in Stated Preferences (SP) surveys, which do not always unveil the same choice patterns as Revealed Preferences (RP) surveys. This issue happens because the former measure desired scenarios and the latter what scenarios took place in reality. Such discrepancies with real-world behavior often challenge the mentioned theoretical assumptions.

1.3 Project objectives

This project's primary objectives and milestones relate to analyzing and modeling the factors impacting mode choices in Asprela. These included identifying and grouping these factors a priori from the data collected via an RP survey conducted as part of the MOSH project in its previous stages. The task involved performing a thorough descriptive analysis of the collected sample and comprehending the literature surrounding such factors and/or models.

After this initial identification, various DCMs aimed to quantify the relative importance of the studied factors and, when applicable, further refine the set of factors worked with to achieve the most optimal model. Such was always performed with the utmost care to detail through a proposed validation process before accomplishing the end purpose, i.e., interpreting the results of estimations yielded from the mentioned models. The described objectives culminated in a contribution to research and knowledge in transportation, economics, and engineering.

1.4 Dissertation structure

The methodology encompassed a comprehensive approach that started by reviewing the literature in parallel with analyzing the survey data. It followed the modeling phase in an iterative model identification process regarding explanatory factors used and types of model configurations deemed adequate.

This dissertation is structured in 5 distinct chapters, as follows. After the present chapter, Chapter 2 focuses on the review of the literature and theoretical background for the analysis. This literature review chapter is intentionally divided into two sections. The first aims to examine the existing research on transportation mode choice—emphasizing discrete choice modeling applications—and the second is mainly geared towards overviewing the theory behind different DCMs. Most of the document comprises *Methodology and Data* and *Results and Discussion* chapters, corresponding to Chapter 3 and Chapter 4, respectively. The former provides a general description of the survey and, most importantly, of the data collected with it; it follows by characterizing the validation heuristic implemented while applying it to determine the most robust imputation technique applicable and finishes with a comprehensive portrayal of the implemented DCMs and respective validation results. The latter, in turn, presents the findings of the study mainly derived from the interpretation of the previously estimated models. Finally, Chapter 5 summarizes the significant findings while outlining possible limitations of the implemented approach and suggesting directions for future investigation.

Chapter 2

Literature review

This literature review explores the applicability of discrete choice modeling and alternative methodologies to research on transportation, urban mobility, and mode choice. Section 2.1 explores the factors influencing mode choice, mostly in urban settings, categorizing them into four different groups and aiming to deepen the understanding of how individual preferences and contextual elements interact to shape traveling behavior. Section 2.2, in turn, provides an overview of DCMs, describing the most common applications and extensions, particularly in their role in mode choice analysis.

2.1 Transportation mode choice

The preeminent interest of this study is to understand the factors that drive these choices in an urban context, better comprehending the traveling behavior and mode choice intricacies of the individual agent present in these surroundings. Therefore, given this nature—and the data utilized—the emphasis is shifted towards studies conducted in such settings. Only with proper knowledge regarding best practices, research gaps, and promising areas of future research can modern societies develop and apply effective transportation policies that can promote the always-aimed sustainability and efficiency in an urban context.

Types of factors analyzed Following partially the framework proposed by Zhou (2012) when reviewing the literature on mode choice, relevant factors expected to influence mode choice are divided into four sections. Given the interchangeability of some factors and that mode-specific factors and trip characteristics could be considered a group of less abstract or more instrumental factors, both groups compose a broader *Objective Factors* set. This interchangeability may also mean that some mode-specific factors, given how they are perceived and experienced, could and should be seen jointly with psychological and experience-related factors under *Subjective Factors*. Disregarding individuals' socio-demographic attributes from their socioeconomic or geographical context is impossible. Hence, these factors are presented as *Socio-Demographic, Economic, and Cultural Factors*. Notwithstanding the connection of this last group of factors to physical location,

an additional *Built Environment* group pertains to the influence of specific physical characteristics where traveling takes place.

2.1.1 Subjective factors

Individual perception, personal traits, and attitudes can all play a significant role in mode choice outside of more straightforward factors like cost or travel time, and even the way these two influence the final decision is subject to the potential existence of boundaries set by perception and intuition. This seems rational since users who travel less frequently are likely to attribute higher monetary value to time savings (Cantillo et al., 2006; Johansson et al., 2006). Additionally, the literature consistently supplies evidence on how past experiences with a specific mode of transportation can influence perception and guide the individual to choose that or another mode (Abou-Zeid et al., 2012; Hensher and Ho, 2016; St-Louis et al., 2014; Whalen et al., 2013).

Perceived quality of service Similarly, perceived service or performance levels and safety (Johansson et al., 2006) of public transportation significantly impact its attractiveness. Including such factors in discrete choice modeling as latent variables results better than traditional approaches. Personal characteristics and lifestyle can also weigh as much the purpose of the trip—e.g., occasional study trip or regular commuting—as shown in a qualitative approach by Beirão and Cabral (2007). Despite using a different methodology and considering the period since their publication, most findings from this study are of special attention for the present research, for it studied individuals from the metropolitan area of Porto. As demonstrated further in Dell’Olio et al. (2011) regarding public transportation, through a series of logit models applied to stated SP surveys, besides perceived quality of service—related to daily experiences of the users—, desired quality of service—what users desire, hope or expect from these modes—should be scrutinized when trying to, through marketing policies, capture potential users, i.e., those do not currently use public transportation and whose most valued factors differ from actual users. The authors state that not being a bus user might be highly significant when judging the factors linked to public transportation’s quality of service.

Satisfaction and perceived first choice of traveling Within the scope of these factors, traveling satisfaction emerges as a prominent element. One can take this factor as an interaction between objective and subjective factors, depending on them but also on the mode of transportation itself. Individuals who frequently prefer non-motorized modes of transportation—pedestrians or cyclists—and train commuters—mode interaction with purpose and regularity—tend to report higher satisfaction levels than car owners and metro or bus users, highlighting benefits they can never achieve with such modes (St-Louis et al., 2014). Nevertheless, distinct modes are associated with distinct satisfaction determinants. For example, external factors such as congestion or delays typically negatively impact metro and bus users. Conversely, the case can also be made for mode preference, i.e., considering a particular mode as the first choice, which conditions these public transportation users the most (St-Louis et al., 2014). Masoumi (2019) also cites personal

preference for car use as the main barrier to public transportation use in the Middle East and North Africa (MENA) region.

Mode characteristics' impact on individual perception Moreover, reiterating the individual's psychological aspect, commuters perceive value beyond simply traveling from point A to point B on time. While some enjoy the traveling experience or relaxation provided, others lean towards using time productively, using the mode's service characteristics—from Wi-Fi availability to the presence of individual entertainment systems. In contrast to previously mentioned factors, this typically holds for all the most common modes (St-Louis et al., 2014; Acker et al., 2020). Anable and Gatersleben (2005) examined affective-related factors in these circumstances, separating trips into commuting and leisure. On the one hand, although instrumental factors, e.g., convenience, were deemed most important overall for commuters, some affective factors, such as feeling in control or not being stressed, still ranked as high as instrumental factors' average importance. On the other hand, leisure trip respondents seem to value both groups of factors equally, with relaxation and a sense of freedom being the highest-ranked factors. These findings also hold significant value for objective factors like convenience and trip purpose, discussed further in Section 2.1.2.

Past experience with different modes The experiment conducted by Abou-Zeid et al. (2012) submitted car commuters to a temporary mode switch to public transportation. It significantly influenced short-term satisfaction and travel happiness, with participants reporting higher satisfaction levels with the commute by car post-treatment. Even though the intervention temporarily altered perceptions and experiences, it naturally did not follow a permanent switch for participants, who appeared to have only confirmed their negative bias against public transportation options. However, interestingly enough, some continued to use such options occasionally. Thus, ignoring past individual experiences with various modes is not a good option. Hensher and Ho (2016) found this to be accurate, especially in the case of discrete choice modeling, stating that models that do not explicitly consider past experiences underestimate the individual's sensitivity to direct changes in trip duration or cost. For public transportation, e.g., with shorter trip durations, the experience-enriched model would reflect a higher market share increase than a standard multinomial logit that does not account for experiences.

Environmental issues Johansson et al. (2006) found that being pro-environmentally inclined significantly influences mode choice, particularly by increasing the likelihood of choosing train over bus. Nevertheless, the factor does not make a difference when choosing between driving and taking a bus. This finding further highlights bias towards car use and the discrepancies between stated preferences and actual behavior. According to Beirão and Cabral (2007), interviewees did not link any environmental concern to car use.

2.1.2 Objective factors

Objective factors can include any directly measurable factor that is not socio-demographic or specific to the individual. In this framework, for the present study, these factors are considered chiefly a mixture of mode-related factors and factors derived from the characteristics of the trip itself.

Cost Cost is extensively considered one of the most critical factors of mode choice in urban settings, and an extensive portion of research has carefully examined its impact (Beirão and Cabral, 2007; Cantillo et al., 2006; Horeni et al., 2007; Salon, 2009; Tyrinopoulos and Antoniou, 2013; Acker et al., 2020). Even when its study is not prioritized, authors cannot disaggregate it from their analysis once its relation, indirect effect, and comparison with other factors is inevitable (Hess et al., 2005; Johansson et al., 2006; Whalen et al., 2013; Anable and Gatersleben, 2005; Cervero, 2002). Although car ownership links significantly to trip cost for those who choose to drive a private vehicle, such factor is analyzed separately as a socio-demographic component in Section 2.1.3.

In alignment with the theories of rational economic behavior, the decision-maker expects cost to represent a *disutility* (Train, 2009d; Hess et al., 2005). Nonetheless, some novel findings shed new light on this issue, as not every individual behaves like a true *homo economicus* (Hess et al., 2005). Beirão and Cabral (2007) demonstrated an interesting choice paradox: while most car users admit bus traveling would be cheaper than driving, they are reluctant to change modes; at the same time, cost itself is perceived as the most damaging aspect of driving. Such might imply that cost is not as important for these individuals. Their inability to gauge how much a car trip costs, almost always underestimating this value by only considering fuel and parking costs, further confirms this. Conversely, Tyrinopoulos and Antoniou (2013) complement these findings, where public transportation users were deemed not discouraged by potentially high fares associated with these modes. In their estimation of an MNL with data from a university setting, Whalen et al. (2013) found that owners of a parking permit are not sensitive to its cost, given that owning it increased their utility. This positive impact on utility arguably happens because they typically pay for the permit as a seasonal ticket, whether using it or not. Anable and Gatersleben (2005) mention the interaction effect of reaching a specific destination—which can reflect the trip purpose—and both cost and trip duration, concluding that individuals do not necessarily minimize their travel time or choose the most cost-efficient mode for specific destinations.

Studying car-dependent individuals commuting in New York, Salon (2009) found that the population is more sensitive to trip duration changes than to cost, though still seeing the latter as a *disutility* for all alternatives, regardless of income level. In dissonance, an experimental study focused on car-use cost fluctuations showed that individuals prefer to reduce travel costs instead of travel time when unable to minimize both. Researchers also observed a shift to other traveling alternatives. An important caveat is that the participants of this study were university students, possibly downplaying the influence of income on the observed trade-off (Horeni et al., 2007). As per Zhou (2012), university students differ in traveling behavior from the general population.

Sometimes, cost is no more impactful for mode choice or held in higher regard than flexibility and comfort (Johansson et al., 2006; Anable and Gatersleben, 2005). Nonetheless, in the realm of long-distance, Acker et al. (2020) found that travel cost, although not to all groups of travelers, is the dominant factor: even when considering similar service levels to what other modes offer, ticket price prevails as the predominant determinant of bus services.

Trip duration Rivaling with cost comes travel time because of both its importance on mode choice and research interest. Indeed, this factor's exploration appears almost a constant across all literature, particularly its intricate relationship with cost (Cantillo et al., 2006; Hensher and Ho, 2016; Horeni et al., 2007). This factor is often considered the most important determinant of mode choice, as in Ton et al. (2019) or Salon (2009). For the former, this conclusion was the most appropriate among all the alternatives studied: public transportation, car, cycling, and walking. The study by Salon (2009), in turn, holds particular interest since the reported adverse effect of trip duration is primarily a reflection of waiting time, whose estimated coefficient is approximately 7-8 times higher than the one computed for the time inside the vehicle. Moreover, waiting time also possesses the most significant positive elasticity impact on car use for the same study. Such findings are congruent with the statement in Dell'Olio et al. (2011) that waiting time, potentially representing time the individual perceives as lost, is often one of the factors with the highest weight in the utility function of a DCM. In this study, waiting time and time inside the vehicle are precisely two of the three most valued factors for actual and potential users.

Cantillo et al. (2006) suggest the integration of perception thresholds in DCMs for trip duration, arguing that time savings, if not sufficiently large, are often imperceptible by the decision-maker. The individual utility, otherwise increased by applying these time savings to other activities of personal interest, is modeled as static for a given time interval. Such economic implications are possibly not as applicable to cost since individuals can reallocate almost any amount of money to generate higher utility. More counterintuitive findings, according to economic theory, are associated with trip duration: in Whalen et al. (2013), researchers recognize a positive utility of time, particularly for bicycle and car, severely more likely to be chosen as trip duration increases in comparison to public transportation. Access time, i.e., the time to reach the transit network's entry point, was also significantly impactful.

Trip duration interacts with trip purpose, as demonstrated by Buehler (2011): in Germany, while shorter trips see significant odds increase of walking over driving, it is also true that similar increases happen to cycling when the trip represents a commute to work, thereby suggesting this is a competitive alternative to driving for short work trips. Conversely, for more extended work trips, the study concludes that individuals are more likely to choose public transportation over car driving. This finding also applies to the USA, according to the same study. Such conclusions are arguably expected, as public services seem more geared towards commuting, mainly in rush hour traffic, being faster than cars for longer trips due to fewer stops.

Trip purpose According to Acker et al. (2020), the value of time, hence the importance attributed to travel time, varies between individuals also from the interaction with socio-demographic characteristics or trip purpose: people with higher incomes, like business travelers, might be willing to spend more money to save time they would, otherwise, be used to travel. Another interesting finding from this mixed logit application is that, in contrast to other groups, it expected leisure travelers to associate higher utility with higher costs, concluding that these individuals, who essentially travel for pleasure, are the least price sensitive. In a study comparing the importance of socio-demographic characteristics of tourists and the motivation behind their trips, Romão and Bi (2021) found that such motivations tend to shape mode choice more strongly than socio-demographic factors. Professional travelers are typically more car-oriented than general, notwithstanding leaning towards public transportation during more extended stays, which also results in higher perceived satisfaction from these users. The importance of travel time to individuals commuting to work also applies to study-related trips (Beirão and Cabral, 2007).

Convenience, flexibility, and comfort The research found that convenience (in the form of a mixture of "catering," "speed," or "Wi-Fi") and comfort (in the form of increased "leg space") play a critical role in long-distance traveling, with the latter only being trumped by cost, as per Acker et al. (2020). Not necessarily measured directly, researchers can infer convenience using proxies, such as the number of modes used to perform a given trip. In this context, according to Beirão and Cabral (2007), individuals prefer a more direct public transportation service since, in general, they dislike having to change vehicles during a journey. The study adds that people pursue a more comfortable and relaxed journey, preferably seated in a less crowded vehicle. This finding aligns with Tyrinopoulos and Antoniou (2013), where the research found that crowding is the most discouraging factor in public transportation. Dell'Olio et al. (2011) also mention the factor as one of the only essential factors for both regular and potential public transportation users. Bus comfort is precisely one of the remaining factors, probably due to its correlation with vehicle occupancy.

Thus, if the comfort that allows a possibility of resting, working, and moving around attracts the individual, their likelihood of choosing bus over car and train over bus increases; conversely, if the preference is more directed towards a more flexible trip in regards to higher autonomy and being able to stop anywhere according to personal necessity, car is a more likely alternative in comparison to bus (Johansson et al., 2006). Anable and Gatersleben (2005) argue for higher flexibility of active modes relative to cars. However, the same does not apply to convenience—the most critical instrumental attribute for car users in the said study. The authors expect both factors to outweigh more affective determinants when driving. Finally, Masoumi (2019) also found that the lack of comfort and convenience generates the highest impact on avoiding public transportation in favor of driving a car.

Trip regularity, mode frequency, and availability The inclusion of trip regularity, mode frequency, and availability under the umbrella of either subjective or objective factors is debatable.

Trip regularity, also known as usage frequency, can be a fundamental determinant for mode choice, particularly for potential users, whose perceptions regarding a particular mode can be substantially different from regular users (Anable and Gatersleben, 2005). Hensher and Ho (2016), e.g., used this factor as a proxy for past experiences. Besides, individuals might be more accustomed to traveling a certain way for a given destination without adequately weighing the benefits of other alternatives.

On-demand availability, deeply linked with the frequency of vehicle passage in the case of public transportation, is intuitively expected to interact with other factors like reliability and waiting time (Beirão and Cabral, 2007). Making public transportation modes travel faster and increasing the service offer of such options concerning frequency is discussed as an option to reduce car ownership and/or use in Salon (2009).

DCMs typically process mode availability as a binary outcome. In many choice scenarios, Ben-Akiva and Boccara (1995) state, this treatment as a deterministic outcome is misleading since availability is essentially a byproduct of latent factors and constraints. In Tyrinopoulos and Antoniou (2013), the lack of public transportation information, arguably seen as a proxy or determinant for perceived availability, does not discourage the use of public transportation. Nighttime commuting, in turn, is significantly influenced by reduced public transportation availability (Kapitza, 2022). Car availability is further discussed under *car ownership* in Section 2.1.3.

Weather conditions Weather conditions can influence the choice of active transportation modes, particularly for non-captive commuters, i.e., those with more than one alternative to take their trip. The issue particularly applies to younger individuals concerning colder temperatures, which they typically perceive as discouraging. These modes, especially cycling, are also affected by wind speed and different forms of precipitation. Cars are gradually less likely to be preferred as temperature increases, though such a finding only applies to colder temperatures. Individuals are not sensitive to temperature when choosing public transportation (Saneinejad et al., 2012). The preference for non-motorized modes remains extensive during winter in the context of the study by Whalen et al. (2013). Ton et al. (2019) go beyond and contradict the general body of literature, stating that weather characteristics are not relevant for active mode choice, at least in the context of their study, with structural, cultural, and geographic explanations presented as to why the Netherlands might be an exception.

2.1.3 Socio-demographic, economic, and cultural factors

As described further in Section 2.2.1 and Section 3.3, overall satisfaction from any choice, i.e., utility, is a byproduct of an individual's specific characteristics. Most commonly, these encompass their socio-demographic information. Additionally, neglecting the inherent relationship between the individual's socio-economic or cultural context and such characteristics is potentially missing part of the explanation as to why they choose as they do. As shown by Masoumi (2019), mode-choice decisions are significantly context-sensitive.

Income level and car ownership Income level and car ownership are some of the most frequently cited factors associated with mode choice across all, not only socio-demographics. This level of interest and knowledge simultaneously is somewhat comprehensible, as both factors are also typically correlated (Salon, 2009). As income rises, owning and maintaining a vehicle becomes more accessible, and the tendency seems to be for individuals to choose the faster available alternative. Since cars are available and the cost of car trips is also relatively lower for higher incomes, this often becomes the most attractive option (Buehler, 2011; Masoumi, 2019; Johansson et al., 2006; Ding et al., 2017; Whalen et al., 2013).

Higher household incomes are, thus, related to higher car traveling and, naturally, less use of public transportation, bicycle, and walking. In his analysis of mode choice in the USA and Germany, Buehler (2011) states that this hypothesis is consistent for both countries. Nevertheless, in the USA, car use still dominates traveling among all income groups, while Germans have a higher reliance on active modes and public transportation, suggesting cultural or policy influence.

The mentioned structural equation modeling (SEM) approach by Ding et al. (2017) critiques the treatment of car ownership as an exogenous variable typically observed in discrete choice modeling. Hence, they modeled the mediating effects of car ownership to infer about built environment characteristics' impact on mobility. According to the authors, there is established consensus about household car ownership as one of the determining factors of mode choice. Buehler (2011) and Ding et al. (2017) further verify this, stating that individuals in households with more cars per household member choose to drive it more frequently.

Even in countries like India, where researchers expected the level of densification to play the most critical role in mode choice, household income is found to be the most crucial determinant of mobility, with higher levels associated with higher shares for private and/or motorized modes (Masoumi, 2019). However, there might be a caveat. In more developed and industrialized countries, where the probability of owning a car is substantially higher, these more economic-inclined factors are arguably less relevant. Consequently, other demographic factors, such as household composition, age, or gender, might have a more significant role in mode choice (Buehler, 2011). Other authors speculate that a shift in car ownership is already observable, potentially accelerated by the advent of several sharing-inclined mobility alternatives (Ho et al., 2018).

Age Younger individuals are more likely to opt for active modes or public transportation, particularly in car-dominant cities, possibly due to lower job or household responsibilities and lower willingness to pay concerning private cars (Zhou, 2012). The same study states that such is generally also valid for university students, potentially reflecting the interaction effect with low/no income and younger age. This finding is arguably consistent with urban mobility in Greece, as individuals aged between 35 and 44 preferred car driving (Tyrinopoulos and Antoniou, 2013). The latent variable model used by Johansson et al. (2006) discloses that age is one of the roots of individual preference discrepancies. Ding et al. (2017) further extend the argument, stating that younger and older individuals are both more likely to choose public transportation or active modes relative to middle-aged people, who, in turn, tend to own more cars and travel longer distances.

Using the reasons to choose (or not) a particular transportation mode, Masoumi (2019) built several MNLs to analyze mode choice in the MENA region, with data from Tehran, Istanbul, and Cairo, concluding that age is significantly related to public transportation use: the older the individual, the less likely he is to use it. The effect also applies to opting for a bicycle.

Gender Regarding gender, the conclusions point to unique directions, depending on context. The responses to the interviews by Beirão and Cabral (2007) suggest that women, in particular, would prefer public transportation over car driving if suitable for them. Tyrinopoulos and Antoniou (2013) and Zhou (2012) reiterate this preference in their findings. The latter adds that women are significantly more likely to go biking or walking. Johansson et al. (2006), in their turn, found that women are more environmentally inclined than men and, consequently, most likely to opt, e.g., for train over bus. Ding et al. (2017) fail to diagnose the influence of this factor with complete confidence, but the results suggest men are more likely to choose to drive a car. The potential inclination of women for non-motorized modes compared to men is, nonetheless, entirely rejected by the logit and nested logit applications of Rodríguez and Joo (2004). Men are also more likely to travel long-distance (Acker et al., 2020; Ding et al., 2017). Nighttime traveling displays complex interaction effects with gender. Potentially resulting in increased insecurity, it is generally associated with a higher likelihood of driving a car (Kapitza, 2022).

Although not statistically inferred, Masoumi (2019) adds that many Muslim societies often possess cultural barriers—including government limitations—against women's cycling, indicating a potential role in their traveling decisions, even if indirectly.

Household composition The traditional discrete choice approach by Johansson et al. (2006) leads to the conclusion that individuals in households with children are more likely to choose to drive a car over traveling by bus. In the same study, however, such was inconclusive in their latent variable enriched model. According to Ding et al. (2017), household size significantly influences mode choice—though as an indirect effect, with bigger size related to owning more cars. Additionally—here as a total effect—the number of children in households resulted in travelers with a lower likelihood of choosing walking, cycling, or public transportation, notwithstanding the conclusion that having more children results in less car ownership. The latter is naturally related to the driving age policy. Beirão and Cabral (2007) also mention car dependence to take children to school as an obstacle to public transportation. Students who live alone are more likely to choose to drive solo (Zhou, 2012).

Education level A less-mentioned factor is education level. Ding et al. (2017) addresses this problem, concluding that individuals with higher education levels lean towards car use. Researchers speculate that such individuals might be involved in longer distances for work and social activities, thus increasing the likelihood of driving. Besides, the authors state that this factor positively influences car ownership for the same household income. Individuals who are students or have completed a college degree have a higher propensity to use bicycle Ton et al. (2019).

Public transportation subscription Lastly, as expected, owning a public transportation subscription was positively related to the likelihood of traveling with such modes. Interestingly, Ton et al. (2019) also found the same effect concerning cycling. In their study, authors suggest that this effect might be due to the not-so-unusual usage of bicycle as a complement for the whole journey.

2.1.4 Built environment

In the domain of mode choice, the built environment refers to the physical characteristics shaped by humanity in certain regions, cities, or even neighborhoods, which intrinsically can influence how people travel. Among them, the most usually quoted are population density, mixed land use, topography, accessibility, availability and quality of infrastructure, mainly relative to public transportation, and urban design (Buehler, 2011; Cervero, 2002; Ding et al., 2017; Rodríguez and Joo, 2004; Tyrinopoulos and Antoniou, 2013; Wardman et al., 2007). Individual case studies provide critical knowledge about the relationship between the local physical environment and travel behavior. Nonetheless, its findings are scope-limited and may not be straightforwardly generalized to a broader picture (Rodríguez and Joo, 2004). On top of that, most of the DCMs estimated to investigate these factors present an incomplete specification, partly because of modeling conventions and partly because of data quality-related limitations (Cervero, 2002).

Population density Concerning population density, the current consensus is that, on one side, more densely populated areas, reporting higher employment densities, are more favorable to public transportation utilization. Conversely, less dense regions shelter individuals with higher car dependence, therefore appearing to display higher degrees of car ownership (Cervero, 2002; Salon, 2009; Ding et al., 2017). This finding is consistent for both American and European countries (Buehler, 2011), even though such an agreement might not indeed hold for all cases. While accepting this theoretical significance of population density, various DCMs, applied to commuter survey data collected in a university context in a study by Rodríguez and Joo (2004), did not consistently relate the factor to mode choice: researchers speculate that employment densities at destinations, rather than population densities at origins may better explain mode choice. It is crucial to understand that population density, by itself, is not the only driving force behind mode choice shifting: it often acts as a surrogate of other vital factors that highly correlate with it, and excluding them from modeling usually leads to its overestimation. A discrete choice analysis in India concluded that public transportation is more prevalent in larger cities with higher densities. In smaller cities, individuals exhibit more car dependence (Ahmad and Puppim de Oliveira, 2016).

Mixed land use Concerning mixed land use, the authors suggest increasing it would result in a higher propensity to choose non-motorized modes and a lower propensity to use motorized modes. A decreased commuting distance and the discrepancy in the valuation of time-travel savings explain this finding: individuals place a higher value on such savings for non-motorized modes than in-vehicle or motorized modes. Other studies (Buehler, 2011; Cervero, 2002) do not fully support this idea, yet advocate that higher land mix-use levels encourage public transportation riding

against drive-alone or car traveling. Contradictory to all these insights and initial expectations, the discrete choice and integrated SEM application, by Ding et al. (2017), unveiled that higher mixed land use is strongly associated with a higher likelihood of car ownership. The logical interpretation depends on context: while, in some cases, the travel distances are shortened—walking, cycling, and public transportation become more appealing—in others, the reduced travel distances lower the overall costs of car riding—making owning one more affordable.

Sidewalk and cycling facilities availability Urban design-specific characteristics, such as sidewalk availability, could be considered more modest in influence, in contrast to the aforementioned factors (Cervero, 2002). Regarding its particular effect on distinct transportation modes, sidewalk availability significantly links with the propensity to use a non-motorized or pedestrian mode (Whalen et al., 2013), specifically to commute or travel to school (Rodríguez and Joo, 2004). Although speculating about the particular conclusion that cyclists could use sidewalks to travel to campus, Rodríguez and Joo (2004) could not infer such an effect was significant. Nonetheless, the findings are particularly intuitive considering that areas that provide for vehicle traffic—by higher road density—are more favorable to driving a car or finding public transportation services, the same way those that provide facilities for walking tend to increase its utility and, as an aftereffect, cycling utility as well (Whalen et al., 2013). The absence or lack of cycling facilities, added to social and cultural problems, may also reduce the likelihood of cycling in non-Western settings, according to Masoumi (2019).

Parking availability Moreover, in conformity with different studies (Whalen et al., 2013; Salon, 2009), parking availability—whether in the form of a parking permit or, indirectly, in the form of home parking cost—has the potential to define car traveling utility. Tyrinopoulos and Antoniou (2013) relied on the estimation of probit and SEMs and found that parking space availability is the main factor conditioning commuters for car preference. This analysis concluded that parking availability might dominate the mode choice of car drivers, for no other factor from a group of related factors to this one entered their final probit choice model. Beirão and Cabral (2007) found that parking availability is crucial for those who commute to areas where such availability is scarce or parking control exists, with some individuals resorting to illegal parking.

Connectivity and accessibility In addition, compact, well-connected cities and neighborhoods with high degrees of accessibility are interrelated with having residents who are less likely to own a car. The effect is verified, perhaps, because an increased street network connectivity provides a better walking or cycling environment for travelers, but also seeing that households have a shorter distance to public transportation the more compact or well-connected the city is (Cervero, 2002; Ding et al., 2017).

Likewise, Buehler (2011), in his MNL approach to intercontinental data, treats accessibility as a measure of household proximity to public transportation using various distance levels and, in congruence with the previously mentioned findings, established that higher household proximity

to public transportation is associated with reduced car usage in both Germany and the USA, accompanying an increased likelihood of using non-automobile modes. However, Americans rely more strongly on cars than Germans, even when living more closely with public transportation. Conversely, though, Tyrinopoulos and Antoniou (2013) inferred that the lack of accessibility to public transportation networks was not a discouraging factor for commuters at all, meaning in the context of their sample, when such network is accessible within reason, other factors like crowding or service unreliability come out on top.

Topography Although not as studied in the current literature on the role of the built environment, local topography can achieve notable significance when considered in discrete choice methodologies. Everything else being the same, from travel time to demographic characteristics, topography potentially correlates with an individual's propensity to walk or cycle. Logically, additional walking or cycling time differences, although highly small due to slightly steeper terrains, lower the odds of choosing both modes. No impact is detected when walking or cycling to access public transportation, which is also unsurprising since substantially steep hills would be needed to find noticeable differences in shorter walking or cycling trips for this purpose (Rodríguez and Joo, 2004). Wardman et al. (2007) propose implementing measures without contemplating the adverse effect of topography on the propensity to cycle, yet recognize such limitations.

2.2 Discrete choice models

Discrete choice models (DCMs) are one of the most conventionally applied methods for mode choice analysis since they can handle high degrees of complexity, predict probabilities associated with any available traveling options, and, in some configurations, capture the individual preference heterogeneity between individual agents—meaning that different people have different tastes, perceptions, and priorities inherent to each factor of choice. Seeing that these models build on random utility theory—which assumes individuals make rational decisions that maximize utility, i.e., that benefit themselves the most—they grant researchers the ability to quantify the relative influence of each factor, leading to more interpretable results about the decision-making process. Its structure is often flexible, allowing it to be presented in distinct configurations according to the researcher's needs, whether he needs to observe peculiar effects and influences on mode choice, unique contexts, or convoluted relationships between factors.

These models also allow for the estimation of crucial behavioral economics indicators, for instance, willingness to pay (WTP) or the value of time in the form of value of travel-time savings (VTTS) (Hess et al., 2005), which are often valuable to evaluate the benefits from proposed policies or investment in transportation and provide a higher accuracy of resource management.

Even though discrete choice modeling offers many advantages, limitations exist and must be recognized. For example, when dealing with data from SP surveys, which are frequently subject to biases, the true complexity of human decision-making behavior might evade itself. Another

standard limitation is the coherence of the theoretical assumptions of these models with such human behavior in reality.

2.2.1 Multinomial logit model

The Multinomial Logit Model (MNL) is a widely implemented model that has been the cornerstone of discrete choice for many decades. In Train (2009d), the author states it is "By far the easiest and most widely used discrete choice model." It builds on random utility theory (Train, 2009f; Walker and Ben-Akiva, 2002). Thus, it is part of the Random Utility Models (RUMs) family. These models are typically used to model particular choice probabilities from an agent choosing among mutually exclusive alternatives, i.e., discrete alternatives at the disaggregate level. The key assumption behind such models is that individual human behavior equals that of the *economic man*. If the individual chooses an alternative, it is because it maximizes their *satisfaction*. This *satisfaction*, i.e., utility, can be modeled as a linear function of the alternative's and/or individual's characteristics—also known as a systematic component or systematic utility—plus a random error term. In its turn, the random term is assumed to be independent and identically distributed (IID) following a Gumbel (Extreme Value Type I) distribution (Train, 2009d). Still, although theoretically consistent with the utility maximization behavior, RUMs can represent a decision-making process in line with other types of behavior. They describe the relationship of the attributes included in the systematic component to a choice outcome without explaining exactly how the individual made a choice (Train, 2009f).

Advantages and estimation The main advantages of the MNL are its simplicity and ability to compute choice probabilities directly from its estimated parameters without resorting to simulation (unlike probit or mixed logit) (Train, 2009c), making it easier to estimate and interpret. This estimation is conventionally performed by Maximum Likelihood Estimation (MLE) (Train, 2009d; Bierlaire, 2003). Section 3.3 contains further reading on MLE. Moreover, as asserted by Train (2009d), MNL can be seen as the ideal model as long as researchers specify it accurately since their goal is to represent utility so well that anything added beyond what the systematic utility function declares is "white noise," i.e., the error term. Thus, if perfectly applied, it is ideal rather than restrictive.

Limitations In MNL's case, the substitution patterns related to the distinct alternatives are somewhat restrictive and often unrealistic due to the Independence of Irrelevant Alternatives (IIA) property (Bhat, 1995; Train, 2009d). This property implies that the ratio of probabilities inherent to two alternatives is independent of introducing a third alternative on the individual's choice set. In Section 3.3, more details about the theoretical foundations of this model and the MNL proposed by the present research will be presented.

Applications and extensions Although recognizing its limitation, researchers have been using the MNL, together with other variants of the RUM, for several applications and areas: transportation (choice of destination, mode, or itinerary), marketing (choice of packaging, store, or product) or energy (choice of appliances or heating equipment), among others (Bierlaire, 2003; Train, 2009c; Ben-Akiva et al., 2002).

Given the advent of more powerful and flexible models in the late twentieth century into the early twenty-first century (Train, 2009c), the MNL is frequently used as a reference or baseline model before considering the implementation of those less limited methodologies. Ben-Akiva et al. (2002) proposes a Hybrid Choice Model (HCM) framework. Bhat (1995) extended logit to intercity mode choice in his Heteroscedastic Extreme Value (HEV) model, providing a more flexible and precise approach to address the IIA property violation. Cantillo et al. (2006) and Axhausen et al. (2001) incorporate non-linearity in the model's utility specification (Train, 2009d). Golob (2003) reviews the adaptability of SEM by introducing latent variables addressing behavioral challenges in transportation. These can be easily integrated into discrete choice modeling to enrich the analysis by capturing a richer spectrum of decision-making processes, such as unobserved heterogeneity, and psychological factors, such as perceptions and attitudes. The following sections briefly describe other extensions and enhancements.

In any case, Hensher and Greene (2003) suggest that, in discrete choice modeling, using this model should be the first step into empirical investigation before relying on the more complex solutions that can, among other things, be substantially more sensitive to data quality. Whether the analyst should choose a simple MNL or any other model primarily depends on the specific nature of their research. Although the mentioned hierarchy of DCMs exists in the theoretical context, it may not hold in several cases (Bhat, 1995). The fact is that choice type, availability of high-quality data, and applicability of behavioral assumptions or restrictions to real-world matters demand prudence.

2.2.2 Mixed logit model

The Mixed Logit Model (MMNL) (McFadden and Train, 2000), also known as the random coefficients model, overcomes most limitations of the standard MNL, particularly the IIA property. Such flexibility allows for unrestricted substitution patterns, in addition to the introduction of random taste variations, which accounts for the unobserved heterogeneity inside the population.

Advantages McFadden and Train (2000) proved that this model can approximate any DCM derived from a RUM framework, making it a universal tool for discrete choice modeling. As an extension of MNL and still within the RUM family, its utility can decompose into deterministic and random components. Nonetheless, unlike MNL, where this random component follows the IID extreme value assumption, MMNL allows the systematic utility's parameters to vary randomly across individuals, thereby accounting for unobserved individual-specific heterogeneity but also, when specified, for alternative-specific random effects—providing further flexibility to account

for correlations between alternatives. This approach is often a more realistic modeling of choice behavior (McFadden and Train, 2000; Hensher and Greene, 2003; Train, 2009e).

The most straightforward—and most implemented—specification of MMNL is that of random coefficients. Choice probability is a weighted average of the MNL formulation evaluated at different coefficient values, with the weights given by the density function of such coefficients. Altogether, one of the virtues of this model is that the distribution of these coefficients is not limited to the normal distribution (Train, 2009e). It was further demonstrated in Hensher and Greene (2003) the model's ability to handle SP and RP data in different transportation contexts. MMNL is also known for its better handling of panel data compared to MNL (Train, 2009d,e; Hensher and Greene, 2003).

Estimation Due to the discrepancies in their underlying assumptions and mathematical structures, probability estimation in MMNL is significantly more computationally demanding than in MNL. Introducing random coefficients complicates the estimation process, requiring solving complex integrals by approximation. Two alternatives are typically applied: Maximum Simulated Likelihood (MSLE) and Method of Simulated Moments (MSM) (McFadden and Train, 2000). While the first directly approximates the likelihood function by simulating random samples from the random coefficients' distribution, the second matches the simulated and observed data moments. Hence, (McFadden and Train, 2000) emphasize the trade-off introduced by these differences: MSLE might offer better statistical precision but at a significantly higher computational cost, which makes MSM preferable in more complex applications. (McFadden and Train, 2000) discuss the standard simulation techniques further and Hensher and Greene (2003) list software alternatives for practical implementation. The Biogeme package is also a viable option (Bierlaire, 2003, 2023).

Limitations and steps ahead The sophistication of MMNL naturally may lead to complexity issues, from specification to estimation. A common puzzling situation for the analyst is to find the best trade-off of heterogeneity representation, whether through the systematic component design or the coefficient distribution selection (Ben-Akiva et al., 2002). Researchers must carefully identify parameters to avoid biased estimations and/or misinference. In addition, although a generic theme that can be extended to other models (Walker and Ben-Akiva, 2002), MMNL requires extremely high-quality data, and failing such requirement may precede biased results. This prerequisite is highly recommended by Hensher and Greene (2003), although the authors also see this as an opportunity to use the model's capabilities best. They likewise comprehensively enumerate ten specific concerns to consider when implementing MMNL. Such issues, added to the complexities derived from the estimation procedure, must be highly regarded, particularly for researchers unfamiliar with advanced simulation techniques.

Notwithstanding the emergence of artificial intelligence embedded in discrete choice applications (see, e.g., Sifringer et al. (2020)) and the outlined challenges, MMNL is usually still considered one of the most promising discrete choice innovations available (Hensher and Greene, 2003;

Greene and Hensher, 2003), mainly when the traditional logit is not coherent with real behavioral.

2.2.3 Latent class model

Latent class is a fundamental concept of discrete choice modeling. Such is the most basic structure that enables the representation of the population's heterogeneity. In this sense, the Latent Class Model (LCM), motivated by the limitations of logit, starts from the assumption that the researcher can split the population into a finite number of classes. Each class displays its own systematic preferences and distinct behaviors. Hence, the researcher can also model it differently from adjacent classes (Greene and Hensher, 2003).

An extension of logit and a special case of mixed logit Therefore, each class can be a different function of parameters and/or coefficient values that best describe these preferences and behaviors. The heterogeneity assumption ends where the class begins: individuals are still assumed to be homogeneous within classes. As such, LCM is a special case of the MMNL but also functions as a "semi-parametric extension of the MNL" (Greene and Hensher, 2003) since it relaxes certain parametric assumptions while retaining a discrete approximation of heterogeneity: latent classes are nothing more than a discrete approximation of the underlying continuous heterogeneity captured in MMNL. However, these classes are not directly observable—for they capture unobservable heterogeneity—nor deterministically assigned. Instead, each individual is assumed to have a probability of belonging to each class.

Estimation The estimation of LCM involves two main components: class membership probabilities and class-specific choice probabilities. Section 3.5 contains a description of these components. What is important to retain is that the researcher can still estimate both of them following MLE, as in logit. Nonetheless, compared to the most traditional model, the estimation of LCM is often a complex optimization problem, although not excessively, according to Greene and Hensher (2003)—as it would, e.g., be in a mixed logit scenario. The authors consider a good starting value of classes crucial in avoiding the problem.

Applications, advantages and limitations Hence, by not obligating explicit assumptions about the arrangement of parameters across individuals (Greene and Hensher, 2003), unlike MNL (deterministic assignment of constant coefficients for all individuals) and MMNL (coefficients following an arbitrary distribution), LCM provides a notably practical way to handle and interpret the heterogeneity of choice behavior. As an alternative formulation to the MNL, the LCM is typically implemented in similar areas to improve the explanatory power of the former, often also achieving better predictive performance, as in Ben-Akiva and Boccara (1995), where the authors generated the commuters' choice set with an LCM, rather than deterministically assigning it. Walker and Ben-Akiva (2002) lists other examples, mainly related to mode choice. Nonetheless, as of the beginning of the century, Greene and Hensher (2003) argued that applications of LCM to discrete choice with many alternatives were uncommon in literature. The same authors present LCM in a

scenario of road type preferences for long-distance travel in New Zealand, providing four alternatives in an SP survey. They compared the LCM approach to MNL and MMNL. Unsurprisingly, in such a setting, MNL was outperformed by the more complex alternatives. MMNL, in turn, was deemed comparable to LCM, with each model presenting its virtues: LCM favors simplicity and interpretation, while MMNL is more flexible and reliable for precise inference of individual preference.

2.2.4 Nested logit model

The Nested Logit Model (NL) (Williams, 1977) is a generalization of the logit model that admits a specific pattern of correlation of the error terms, i.e., among alternatives. As cited earlier, when not all sources of correlation are explicitly specified, the IIA property does not hold. The NL accounts for such correlation by grouping alternatives into *nests* and is a special case of a second family of models, a subset inside the models that follow the RUM framework, as demonstrated by several authors. These are the Generalized Extreme Value (GEV) models (McFadden, 1977; Bierlaire, 2003; Train, 2009b).

In contrast to MNL, NL includes additional parameters beyond those related to the observed attributes (Bierlaire, 2003). Each parameter is known as *scale parameter* and governs the correlation structure, i.e., the degree of independence of alternatives within or across nests. These scale parameters are interpreted as follows, in line with random utility theory: high values (> 1) translate to low correlation (high independence) and are typically not accepted since violate the theoretical assumptions of the model; low values (< 1) indicate high correlation (low independence); when equal to one ($= 1$), the correlation among alternatives in the same nest is null, and the model reduces to a standard logit if such happens for all nests. In a sense, this is the key feature of NL, which is crucial to distinguish it from the MNL. An important variable not estimated by the model but derived from these dissimilarity values is the *inclusive value* or *log-sum variable*. It summarizes the combined utility of all alternatives within a nest, serving as an explanatory variable bridging lower-level to upper-level choice in the NL model hierarchy. Thus, it measures the overall attractiveness of a nest relative to other nests (Train, 2009b).

Independence of irrelevant alternatives and advantages over logit This model is most appropriate when the decision-maker's choice set is allowed to be divided into subgroups (nests) if and only if, for any pair of supposed homogeneous alternatives inside the same nest, the ratio of their respective choice probabilities is independent of attributes or the existence of additional alternatives. This characteristic partially relaxes the IIA assumption (Bhat, 1995). In other words, as stated by (Train, 2009b), IIA holds inside each nest, but not generally: this is often known as independence from irrelevant nests (IIN). Nonetheless, its structure is relatively simple compared to other models from the same family. It is more complex than the traditional logit and already allows for the mentioned substitution patterns among alternatives. In particular contexts, three- or higher-level NLs—resulting from partitioning each nest into *subnests*—are feasible (Train, 2009b).

Estimation and advantages over mixed logit Regarding estimation, NL resembles MNL as a closed-form model and shares with it the same standard estimation procedure (Train, 2009b; Hensher and Greene, 2003). This attribute provides NL with an outstanding advantage over MMNL, as both typically confront the restrictions of IIA. However, these models are not mutually exclusive, and a *Mixed Nested Logit* is a possibility, thoroughly described in Train (2009g). A practical estimation alternative for NL, although not as precise, is to perform a sequential estimation: first, the researcher must estimate the logits within nests; then, the logit of choice among the nests. Nonetheless, MLE still governs each step. Train (2009b) presents this process most effectively.

Applications and extensions Being possibly the most popular GEV model after MNL (Bierlaire, 2003; Train, 2009b), it has been, similarly to the latter, applied to a great variety of fields—with a significant focus on transportation—, often also as a reference model for theoretical inference. In Axhausen et al. (2001), the authors extended the comparison to the most common members of the GEV family, including the cross-nested logit (further reading in Section 2.2.5). Wen and Koppelman (2001) extended the model by relaxing constraints on the scale parameters in their study on mode choice behavior. More recently, Sifringer et al. (2020) boosted the NL by integrating neural networks to represent non-linear relationships of the observed data, yielding superior prediction ability and estimation results compared to the traditional approach. They introduced it as the Learning Nested Logit (L-NL). The same extension was applied to MNL, resulting in the Learning Multinomial Logit (L-MNL).

2.2.5 Cross-nested logit model

The Cross-Nested Logit (CNL) model, first specifically mentioned by Vovsha (1997) in a mode choice application inside a metropolitan area, is a generalization of the NL. By accommodating overlapping nests, where alternatives can belong to multiple nests with varying degrees of membership, the model allows for substitution patterns of higher complexity level (Wen and Koppelman, 2001; Bierlaire, 2001).

Flexibility, estimation and other advantages over traditional methods The model's flexibility is beneficial in applications where alternatives cannot, a priori, be mutually exclusively grouped in homogeneous groups under IIA (Wen and Koppelman, 2001), enabling mixed interactions and a more nuanced understanding of consumer behavior. Such was one of the frequent critiques of the NL denoted by Bierlaire (2001). Although relaxing the rigid hierarchical assumptions of NL, it still complies with the GEV conditions (Bierlaire, 2003), being one of the most popular integrants of this family of models after MNL and NL. CNL also preserves closed-form choice probabilities, and their formulation—read Wen and Koppelman (2001), where the authors present the restriction differentiating CNL from the introduced generalized nested logit—reflects the shared contributions of the said overlapping nests.

Applications The case study by Axhausen et al. (2001) involving preferences regarding a new high-speed rail system highlights the comparison of CNL to other models of the GEV family estimated with the Biogeme package (Bierlaire, 2003, 2023)—said being able to estimate any GEV model. The evident conclusion was that higher complexity in model structure leads to an improved quality of the estimated coefficients: NL outperforms MNL, yet CNL outperforms it.

Limitations Nonetheless, as pointed out by the authors, such inference must be cautiously generalized, as the application dealt with few possibilities of nest structure. High flexibility might have drawbacks. In the context of several alternatives, this number of structure possibilities promptly increases, leading to over-parametrization in some cases and hindering the search for the best structure solution in others—or both simultaneously. The analyst must often hypothesize and judge which options to experiment with to overcome the issue (Wen and Koppelman, 2001). Estimation can also be seen as a limitation in some cases, as the model may become overly complex for the former reasons (Wen and Koppelman, 2001; Axhausen et al., 2001). This concern also applies to the NL, although not as gravely Bhat (1995).

Chapter 3

Methodology and data

The discrete choice modeling approach is the framework adopted by this study to analyze the factors that shape transportation mode choice in Asprela. Grounded in random utility theory, these models assume that decision-makers choose the alternative that maximizes their perceived *utility* (Y) from a finite set of alternatives. In the present context, the decision-maker may choose to travel by bus, metro, taxi, ride-hailing platform, own car, riding with someone, bicycle/scooter, or walking, depending on their perceived availability of such alternatives. This concept of *utility* is a product of observed *factors* (X_k) weighed by the respective *coefficients* or *parameters* (β_k). Mathematically, it can be expressed by none other than the familiar multiple linear regression formulation:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon \quad (3.1)$$

To better understand the structure of the present chapter, one must pay attention to the fact that the description of the explored implementation is not strictly sequential. This chapter begins by briefly overviewing the survey that allowed collecting the working data, as it does with the data itself, in Section 3.1. This section summarizes the initially selected *factors* for the implemented DCMs that lay the foundation for this study. Such contextual basis is essential to understand the *modeling phase* that follows.

The primary purpose of this chapter is to describe the process through which this study fulfilled its goal, which was finding the best *combinations* of *utility specification* (i.e., the set of explanatory factors) and *model structure* (e.g., MNL or NL) that yielded the best-performing DCMs for the present context. For such *combinations* to be as optimal as feasible, the study proposes the development and employment of a model *validation heuristic* in Section 3.2. This *validation heuristic* culminated in an iterative framework implemented anytime a *parameter estimation* of any DCM occurred, intending to assess either the model's quality or the relevance of the *factors* on which it builds.

Section 3.2 presents the *validation heuristic* implementation as a preliminary step to understand the best missing data imputation technique for the present application. Such allows a simple illustration of the rationale behind and how it works in practice.

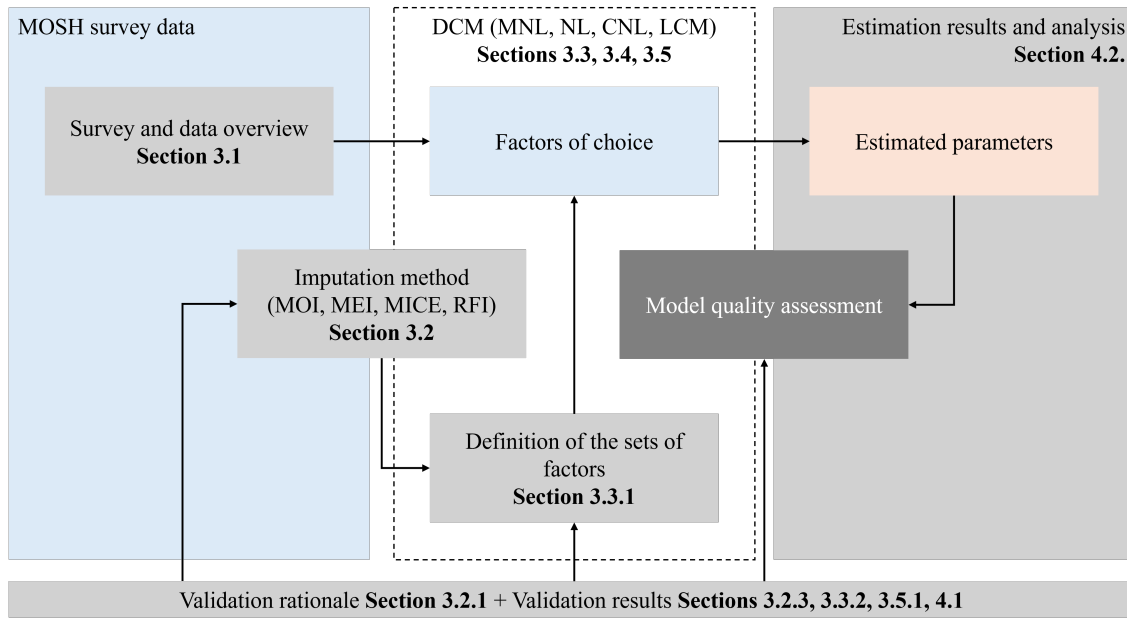


Figure 3.1: Methodology and discussion outline.

Nonetheless, to implement any DCM, the study needed, in parallel, to address in detail how these models are built, i.e., what is the typical estimation procedure, how the model should be specified, what assumptions are allowed, and what factors should be included or excluded after any estimation. Section 3.3 describes how to do this, using the multinomial logit model (MNL) as a reference to address the mentioned topics. In reality, it is noteworthy that, given its iterative nature, the *model validation heuristic* in Section 3.2 uses precisely the concepts only outlined later in Section 3.3.

In its final Sections 3.4 and 3.5, this chapter explores more sophisticated models to extend and complement MNL, mainly tackling the independence of irrelevant alternatives (IIA) property. The said best *combinations* are then studied in the next chapter, by mainly interpreting the estimated *parameters* β_k that fit the models to the working data. A full outline of the integrated structure of the proposed methodology and discussion is shown in Figure 3.1.

3.1 Survey overview

As part of the MOSH project, this study used several sections of the survey proposed by Mahani et al. (2025) in their work on stated and revealed preferences of individuals traveling in Asprela, Porto. The survey took place between November 20–24, 2023, and resulted, after thorough quality control, in a total of 698 valid answers. For greater detail and in-depth understanding of the data collection methodology, the present study recommends further reading of Mahani et al. (2025), as the focus of this section leans more towards the data itself.

Thus, the following descriptions follow the survey’s nomenclature and formulation. To bridge the gap between the two types of preferences, the proposed methodology in this study includes the

Table 3.1: Breakdown of MOSH project's survey. Only revealed preferences sections were considered.

Factor / Variable	Group	Section	Number	Question / Statement
Age	SD	Screening	1	How old are you?
Trip character	OF	2A	35	Is this trip a regular or occasional trip?
Trip regularity	OF	2A	36	How often do you take this regular trip?
Trip purpose	OF	2A	37	What is the objective of this trip?
Trip direction	OF	2A	38	Is this trip part of an outward or return journey?
Starting time interval	OF	2A	39	What time did you start this trip?
Availability	NA	2A	40	What transportation mode(s) were available to you for this trip?
Choice	NA	2A	41	What transportation mode did you choose for this trip?
Number of modes	OF	2A	42	During your entire journey, how many different modes of transportation did you use besides walking?
Trip duration	OF	2B	43	How long did this trip take?
Total duration	OF	2B	44	How long did the total journey take?
Cost	OF	2B	45	How much did this trip cost?
Walking time (bicycle/scooter)	OF	2B	46	How long did you walk to the location of the bicycle or scooter you used for this trip?
Walking time (own car)	OF	2B	47	How long did you walk to the vehicle?
Walking time (bus/metro)	OF	2B	48	How long did you walk to the bus/metro stop?
Waiting time (bus/metro)	OF	2B	49	How long did you wait for the bus/metro stop?
Waiting time (taxi/ride-hailing)	OF	2B	50	How long did you have to wait for your vehicle?
Walking time after drop-off	OF	2B	51	How long did you walk from the place where the mode of transport used dropped you off to your final destination?
Enjoyment	SF	2C	52	I have truly enjoyed traveling this way / with this transportation mode.
Perceived wisdom of choice	SF	2C	53	My choice to make this trip was a wise one.
Satisfaction	SF	2C	54	I am satisfied with my most recent decision to travel this way / with this transportation mode.
Happiness	SF	2C	55	I am happy I made my most recent trip this way / with this transportation mode.
Affordability	SF	2C	56.1	The price of this mode of transportation (how economical it is).
Convenience	SF	2C	56.2	The overall convenience of using this mode of transportation.
Feeling of control	SF	2C	56.3	The extent to which this mode of transportation gives you a feeling of being in control.
Cost-benefit	SF	2C	56.4	The overall value you get from this mode of transportation for your money and effort.
Willingness to travel again	SF	2C	57	I will travel more with this transportation mode in the next few weeks.
Positive things to say	SF	2C	58	I say positive things about this transportation mode to other people.
Perception as first choice	SF	2C	59	I consider this transportation mode to be my first choice to travel to my destination.
Weather impact	SF	2C	60	To what extent did the weather affect your travel experience today?
Sex	SD	3	61	Sex.
Education level	SD	3	62	What is your highest education?
Household composition	SD	3	63	How many people are younger than 13 years in your household?
Disability or impairment	SD	3	64	Do you face any challenges in using public transportation due to a disability or impairment?
Owning a subscription	SD	3	65	Do you have a monthly subscription to public transportation? If so, how much does it cost?
Income level	SD	3	66	What is your approximate monthly income after taxes?

SD represents *socio-demographic factor*, OF represents *objective factor*, SF represents *subjective factor*, and NA represents *not applicable*.

2A represents *Transportation mode choice*, 2B represents *Objective outcomes of the trip*, 2C represents *Experience outcomes of the trip*, and 3 represents *Socio-demographics*. In questions 52-55 and 57-60, the survey asked participants: "Please state your level of agreement regarding the following sentences."

In question 56, the survey asked participants: "Please rate this mode of transportation using a scale of 1 (poor) to 10 (excellent) in the following parameters."

Table 3.2: Results of Pearson's chi-squared test of independence using choice as the target variable.

Factor	χ^2 statistic	p-value
Owning a subscription	253.57***	0.000
Affordability	282.44***	0.000
Trip purpose	237.84***	0.000
Total duration	161.40***	0.000
Feeling of control	228.20***	0.000
Perception as first choice	124.13***	0.000
Trip regularity	131.24***	0.000
Cost-benefit	185.45***	0.000
Willingness to travel again	121.73***	0.000
Enjoyment	120.27***	0.000
Sex	83.43***	0.000
Happiness	95.98***	0.000
Trip duration	97.68***	0.000
Convenience	141.28***	0.000
Weather impact	74.31***	0.000
Income level	96.38***	0.000
Trip character	48.18***	0.000
Satisfaction	72.91***	0.000
Perceived wisdom of choice	70.61***	0.000
Starting time interval	76.38***	0.006
Positive things to say	51.05	0.113
Trip direction	16.54	0.416
Education level	55.50	0.494
Disability or impairment	11.09	0.988
Household composition	20.99	0.994

*** $p < 0.010$; ** $p < 0.050$; * $p < 0.100$

Availability variables and factors non-applicable for all alternatives are excluded.

sections corresponding to trips participants had just taken before completing the survey. For practicality and coherence with the literature review, this study further proposes joining the selected factors into three groups: factors related to *Transportation mode choice* and *Objective outcomes of the trip* collapse into *Objective factors*; *Experience outcomes of the trip* are renamed to *Subjective factors*; *Socio-demographics* keep the same name. In these revealed preferences sections, the researchers asked explicitly about the part of the journey that brought the participants to them, henceforth denoted explicitly as *trip*. Assume, e.g., the following four-trip journey: someone walked to the metro station (1), took the metro (2), then left it and walked to the bus station (3), taking the bus (4) and stepping out where the researchers were conducting the survey. The *trip* the survey studies is the last part, corresponding to the bus (4). Table 3.1. illustrates the questions and respective factors/variables present in the said sections.

Implemented techniques for analysis To better understand the data worked with and, *a priori*, try to grasp how (and if) each factor would relate to transportation mode choice, the study employed the following techniques: Pearson's chi-squared test of independence (using choice as target variable), Pearson's correlation coefficient analysis, and data visualization. The first was

Table 3.3: Results of Pearson's correlation coefficients for non-choice pairs.

Variable / Factor 1	Variable / Factor 2	Coefficient
Satisfaction	Happiness	0.824**
Happiness	Enjoyment	0.678*
Cost-benefit	Affordability	0.672*
Age	Trip purpose: study	-0.645*
Income level	Trip purpose: study	-0.631*
Enjoyment	Satisfaction	0.577*
Perceived wisdom of choice	Satisfaction	0.559*
Age	Income level	0.552*
Walking availability	Number of modes	-0.547*
Total duration	Trip duration	0.542*
Willingness to travel again	Perception as first choice	0.503*
Taxi availability	Ride-hailing availability	0.502*
Trip regularity	Trip purpose: hospital visit	-0.498*
Trip purpose: hospital visit	Age	0.481*
Age	Education level	-0.478*
Total duration	Walking available	-0.457*
Trip character: regular	Trip purpose: hospital visit	-0.449*
Trip character: regular	Trip purpose: study	0.443*
Perceived wisdom of choice	Happiness	0.436*
Income level	Trip purpose: work	0.427*
Cost	Affordability	-0.425*
Convenience	Cost-benefit	0.423*
Trip direction: return	Starting time interval	0.406*
Enjoyment	Positive things to say	0.405*
Convenience	Control	0.395*
Enjoyment	Perceived wisdom of choice	0.385*

** Strong linear relationship; * Moderate linear relationship

The results are filtered from the top 50 highest absolute coefficients.

Redundant and biased correlation pairs and correlation pairs inside the same factor are excluded.

Table 3.4: Results of Pearson's correlation coefficients for choice-only pairs.

Variable / Factor 1	Variable / Factor 2	Correlation
Walk	Trip duration / Total duration	-0.427*
Own car	Cost	0.403*
Own car	Owning a subscription	-0.367*
Metro	Number of modes	0.364*
Own car	Affordability	-0.344*
Walk	Affordability	0.277
Feeling of control	Walk	0.272
Walk	Trip purpose: leisure	0.262
Bus	Feeling of control	-0.260
Trip purpose: hospital visit	Own car	0.258
Cost-benefit	Walk	0.256
Own car	Trip purpose: study	-0.232
Bus	Owning a subscription	0.230
Trip character: regular	Own car	-0.225
Own car	Feeling of control	0.221
Perception as first choice	Ride-hailing	-0.210
Bus	Convenience	-0.209
Metro	Owning a subscription	0.209
Walk	Owning a subscription	-0.207
Cost	Ride-hailing	0.205
Walk	Trip purpose: shopping	0.200
Income level	Metro	-0.199
Own car	Enjoyment	0.198
Own car	Trip regularity	-0.198
Metro	Trip purpose: study	0.197
Metro	Age	-0.183
Ride-hailing	Willingness to travel again	-0.182
Income level	Own car	0.181
Cost-benefit	Own car	-0.180
Happiness	Own car	0.178
Bus	Enjoyment	-0.177
Taxi	Trip purpose: family-related business	0.171
Positive things to say	Walk	0.168
Own car	Age	0.159
Metro	Cost	-0.156
Enjoyment	Walk	0.155

** Strong linear relationship; * Moderate linear relationship

The results are filtered from the top 50 highest absolute coefficients.

Redundant and biased correlation pairs, choice-to-choice pairs, and availability variables are excluded.

implemented as exemplified by Berrett and Samworth (2021) at a 5% significance level (see the detailed results in Table 3.2). For the correlation analysis, its calculation and interpretation are as shown in Ratner (2009) (see the detailed results in Tables 3.3 and 3.4). Lastly, regarding data visualization, the generated plots mainly inspect the differences in the answers to each question across the chosen modes. Nonetheless, for certain factors, even though some of the mentioned techniques suggested weak to no relation with mode choice, a case could be made for their inclusion in the implemented DCMs, whether by inference from the literature or behavioral hypotheses.

3.1.1 Socio-demographics

Questions regarding socio-demographic factors were part of the last homonymous section of the survey, except for the participant's age, which was the screening question used to constrain such age to 18 years or higher. Here, the questions involved the following factors: sex, education level, household composition (i.e., how many people younger than 13 years lived in the respective household), having a disability or impairment, owning a type of monthly public transportation subscription, and, lastly, monthly income level.

Age The sample seems more skewed towards younger individuals, as illustrated in Figure A.1 of Appendix A. The 18-30 year-old group represents $\sim 60\%$ of the population, reflecting their relevance in the context of the Asprela university campus. Moreover, while younger individuals seem to gravitate towards shared and/or public transportation modes, older ones seem to favor own car use and walking, as seen in Figure A.2 of Appendix A.

Sex Figure A.3 of Appendix A shows that, in terms of participants' sex, the sample is exceptionally well-balanced, with approximately 50-50 ratio of male-female participation.

Education level A skew towards higher levels of education reaffirms previous insights regarding the nature of participants and suggests careful analysis and interpretation. The population mainly comprises individuals who finished secondary education or own a degree, as depicted in Figure A.4 of Appendix A. A substantial portion of these expectedly represents university students.

Household composition Figure A.5 of Appendix A illustrates that households with no children below 13 years old are the most frequent inside the sample. By intuition, this reasserts that such a sample was extracted from a younger population and, in general, with many students.

Disability or impairment Only a few participants report having a disability or impairment that challenges the use of public transportation. Interestingly, when distributing the answers by choice, pedestrians have the highest *Yes* percentage to this question ($\sim 5.1\%$), compared to the average $\sim 2.6\%$ of public transportation users.

Owning a subscription Most public transportation users own subscriptions, particularly the most expensive options. This pattern starkly contrasts with the one for car owners and pedestrians, as demonstrated in Figure A.6 of Appendix A.

Income level Among those who report no income or less than 500€, ~80% used the trip to commute to study. Lowest-income individuals, hence most likely students, opt for public transportation, hinting at a heavier dependence on these modes. Figure A.7 of Appendix A shows that these individuals also have a noticeable preference for walking and riding with someone, though not as exacerbated.

Overall criteria for inclusion/exclusion As reported in Table 3.2, the chi-squared test showed no significant relationship between education level and having a disability or impairment with mode choice. The study dropped these factors automatically from further analysis. Although household composition showed similar results, its analysis was deemed relevant, and the proposed methodology kept the factor for in-depth investigation as a binary variable. As shown in Table 3.1, the question corresponding to this factor was "How many people are younger than 13 years in your household?". In this case, the data preprocessing reflects something slightly different, i.e., "Do you have any people younger than 13 years in your household?" with the variable activated (equal to 1) for non-null answers. Owning a subscription showed promising signs of influence on mode choice from all the implemented analyses—similarly to income level—and was also kept to build the proposed DMCs. Even though correlation analysis showed weak to no choice correlation with sex and a moderate positive correlation of age with income level (redundancy risk), as illustrated in Table 3.4, all three factors were considered necessary enough to model. The study further used the binning principle previously applied to household composition to owning a subscription (*yes* or *no*) and income level (*none*, *low*, and *high*). Income level *none* corresponds to *No income*; *low* contemplates the two immediate categories, *Less than 500€* and *Between 500€ and 1000€*; *high* applies to the remaining options.

3.1.2 Transportation mode choice

The first section of questions related to revealed preferences included questions related to the following factors: the character of the trip (i.e., whether the trip was regular or occasional), if regular, the level of regularity, trip purpose (e.g., work or hospital visit), trip direction (whether it was part of an outward or return journey), starting time interval and number of modes used for the respective journey. Beyond that, and most importantly, the survey inquired about which transportation modes were available for the participant for that given trip and what choice they made between the following alternatives: *bus*, *metro*, *taxi*, *ride-hailing platform* (e.g., Uber), *own car*, *riding with someone*, *bicycle/scooter*, *none – I walked the entire trip* (denoted in this study simply as *walking*), and *other*. Since the alternative *other* has no inherent meaning and its interpretation is nearly impossible, it was further dropped from the working sample, reducing it to 687 valid answers with eight possible transportation mode alternatives.

Trip character Individuals use public transportation services predominantly regularly, while private modes see more balance between regular and occasional usage. The latter applies to own car, as illustrated in Figure A.8 of Appendix A.

Trip regularity Figure A.9 of Appendix A demonstrated that most regular users report a high frequency of travel, particularly those who travel with public transportation.

Trip purpose Two-thirds of participants traveled for study and work. These are followed by hospital visitors and leisure travelers, as outlined in Figure A.10 of Appendix A. Together, the four purposes add up to ~90% of answers. Such relevance for study and hospital visits underscores the built environment that characterizes the region from which the data was collected, with Centro Hospitalar Universitário de São João (major hospital and university complex) as its hub. Figure A.11 of Appendix A further shows that individuals used public transportation primarily for work and study while choosing walking and car mostly for leisure and hospital visits.

Trip direction When evaluating if the taken trip was part of an outward or return journey, the data favor outward trips across all modes. This type of trip represents ~61% of the collected sample.

Starting time interval Most participants stated that their trips had started at 7:30-9:30 and 17:01-20:00 intervals, accounting for ~43% of the sample. The distribution of answers to this question is somewhat similar across all modes.

Number of modes As evidenced in Figure A.12 of Appendix A, metro is associated with the highest average number of modes used in combination for a whole journey. Hence, it might play a more crucial role in multi-modal traveling. In sheer contrast, own car is potentially always used as a standalone choice, highlighting a higher convenience for more direct trips. A data issue arises in this regard, as some car owners seemed to have misunderstood the question, answering that they did not use any mode besides walking. This results in an average number of modes slightly below 1.00, which would otherwise be the minimum possible value. The overall average is also noticeably low (1.06). Across all modes, excluding walking, single-mode traveling dominates responses, again with car owners showcasing this behavior the most, as seen in Figure A.13 of Appendix A.

Mode availability and choice Figure 3.2 shows that public transportation alternatives are the most available and chosen alternatives. Individuals typically select walking more times when available, unlike ride services, which seem severely unfavored. Taking a closer look at the most common availability combinations illustrated in Figure 3.3, walking appears as regularly preferred over public transportation as long as it is available. The same occurs with own car, when available

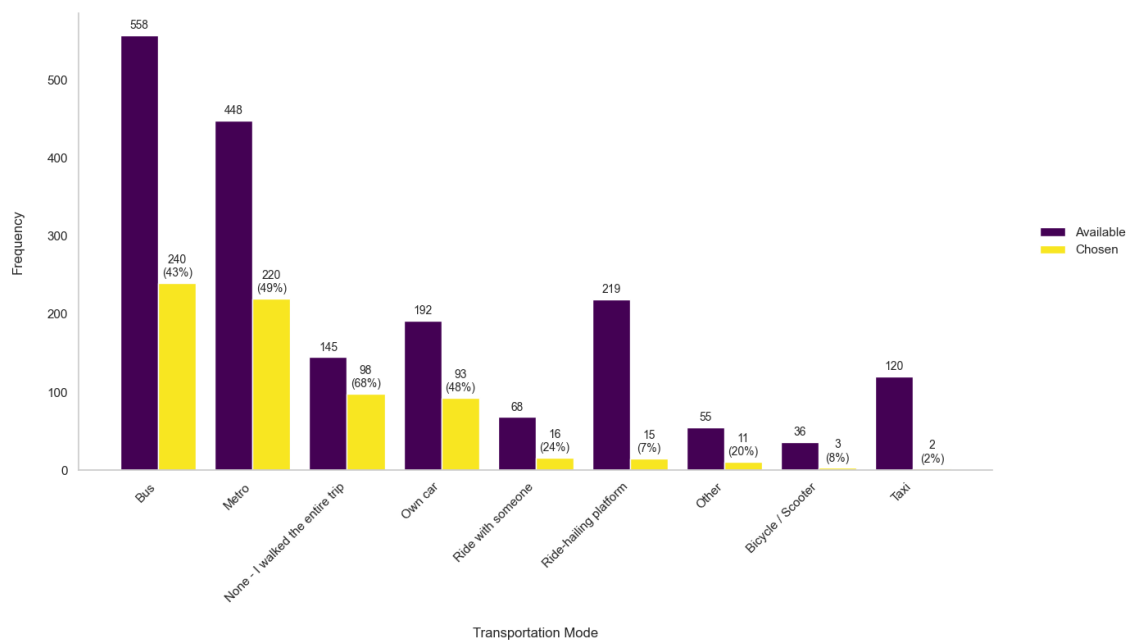


Figure 3.2: Distribution of mode availability vs. mode choice.

with bus and metro. Additionally, many individuals have only one available alternative, making them captive commuters. This scenario strongly affects bus users.

Overall criteria for inclusion/exclusion The results presented in Table 3.2 indicate that the chi-squared test showed no statistical significance in the relationship of mode choice with trip direction and starting time interval. These factors were, therefore, dropped. The choice variable and availability dummies, being fundamental for the construction of the DCMs, were kept, with the same happening to the remaining factors, for they showed all-around positive signs of relation to mode choice. Given that trip purpose is a factor associated with nominal data, each of its unique categories was processed into a dummy variable.

3.1.3 Objective outcomes of the trip

This section of the survey delved into details related to the trip itself. The questions contemplate the following factors: trip duration, total (journey) duration, cost (when owning a monthly subscription for public transportation, the amount besides that), walking time to the vehicle, waiting time for the vehicle, and walking time from drop-off to final destination.

Trip and journey durations Car alternatives display the highest trip durations. Hence, car driving might link with longer-distance trips or trips that ultimately bring the individual closer to the destination, reaffirming its convenience. In contrast, walking trips are likely shorter and more manageable. Public transportation and ride-hailing seem efficient for moderate to long trips, with a good mix of responses—though most options correspond to longer durations (see Figure A.14

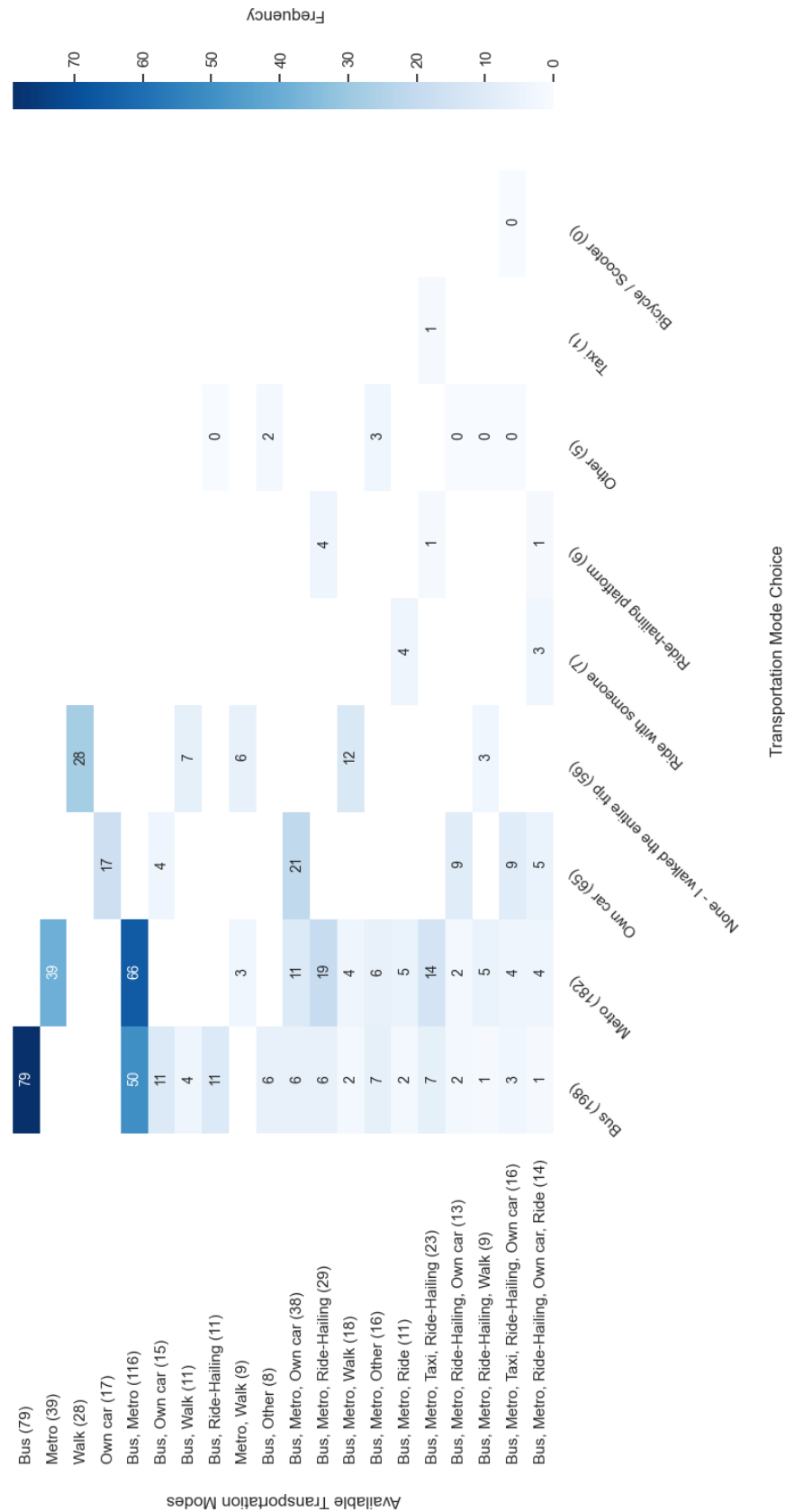


Figure 3.3: Distribution of the 20 most frequent mode availability combinations vs. mode choice.

of Appendix A). However, looking at total duration, car alternatives and public transport become similar, likely due to higher walking/waiting times or multiple modes associated more with the latter (see Figure A.15 of Appendix A).

Cost Public transportation is cost-efficient, while all car alternatives arguably serve users more willing to pay. Car owners and individuals who ride with someone, though to a lesser extent, have a highly pronounced difficulty estimating how much they paid for their trip, as seen in Figure A.16 of Appendix A.

Walking time Figure A.17 of Appendix A highlights that private modes, including bicycle/scooter, require minimal additional walking to reach the vehicle. On average, public transportation naturally presents higher walking times, with virtually no differences between metro and bus. Likewise, on average, car owners walk approximately half as much as public transportation users.

Waiting time Moreover, public transportation users also report the most additional waiting time, particularly when compared to private modes, as indicated by Figure A.18 of Appendix A. The latter are mostly considered waiting-free—except for ride services—with the question not even applicable to them. Nonetheless, the difference is highly more meaningful regarding bus. On average, bus, metro, and ride-hailing users wait 7.26, 4.35, and 4.07 minutes for their vehicle.

Walking time after drop-off On average, public transportation presents the highest walking times after drop-off until reaching the final destination, followed by own car. The differences are, nonetheless, distinctly lower than the ones observed concerning walking time until reaching the vehicle, which might downplay its impact on mode choice.

Numerical approximation The information in this survey section could be converted from ordinal data (numerical intervals) to discrete numerical data. Hence, it was approximated to the said discrete values by assuming a normal distribution inside each interval. An example best illustrates the procedure: if the participant chose the option affirming their trip took *5 to 10 minutes*, the answer is converted to 7.5 minutes. However, in some cases, such an estimate had to be more arbitrary, for lack of defined range: if the option was *More than 15 minutes* for the same question, the answer is converted to 17.5 minutes; for trip regularity, for instance, if the participant chose *3 or more times/week*, the answer is converted to 4 times per week. This study also used the exact approach for instances of this data type in different sections.

This approximation is recognized as a limitation of the data. However, the approach followed is believed to be the best option to preprocess these factors for the modeling phase since the straightforward alternative to such numerical approximation consisted of creating a dummy variable for each bin inside each variable, which would otherwise result in over-parametrization of the model. One could argue that such a dimensionality problem could be mitigated through further binning and that advantages could accrue from it—more straightforward interpretability and

higher flexibility in capturing non-linear effects—yet the chosen solution grants higher efficiency through a more compact representation, making the optimization process more stable. This opposing technique was, in fact, the methodology used in all the aforementioned exceptions: household composition, owning a subscription, and income level.

Overall criteria for inclusion/exclusion From this group of factors, trip duration and total duration showed an upper-moderate positive correlation, as seen in Table 3.4, which is intuitive. To avoid redundancy between factors inside the model, and since trip duration is more important for the context of the model—more coherent with the rest of the data about the trip (not the journey)—total duration was dropped. Walking time from drop-off to final destination was dropped as well: on the one side, it showed not even a weak correlation to choice; on the other side, its relationship with choice in the context of the chi-squared test was biased by those who chose to walk since the question was not applicable for them. Data visualization and behavioral hypotheses also do not support its inclusion in the modeling phase. For the remaining factors, the analyses—in congruence with the literature—showed no good reasons for exclusion (by the contrary). Trip duration and cost appear as the most correlated factors with walking and own car, respectively, as demonstrated by Table 3.4. Likewise, data visualization showed interesting exploratory insights for the said factors and waiting and walking times, whose examination with the other approaches was impossible due to the bias introduced by choice.

3.1.4 Experience outcomes of the trip

Finally, the *Experience outcomes of the trip* section was used to extract the data most related to subjective factors, as described in Section 2.1.1. As before, this experience relates to the trip the participant took. Most of its questions used a 5-point Likert scale (*Fully Disagree* to *Fully Agree*) to assess participants' level of agreement with specific sentences regarding the chosen mode related to the following topics: enjoyment, perceived wisdom of choice, satisfaction, happiness, willingness to travel again, saying positive things about it, perception as first choice (for the specific destination of that trip), and weather impact on enjoyment. Moreover, and still focusing on the chosen modes, it was asked of participants to rate them using a scale from 1 to 10 (*poor* to *excellent*) on these four parameters: perceived affordability (how economical it is), convenience, feeling of control, and overall value in terms of money and effort—denoted further as *cost-benefit*.

The answers to the first group of questions showed slight variation from mode to mode, except for bicycle/scooter. Overall, the differences between questions were also not as significant at first glance. Individuals appear to present a pattern of general agreement with all sentences in this survey section. For such reasons, selecting only a handful of factors and creating a sort of *profile* for each mode was more helpful, which was possible by employing spider charts, also known as radar charts. The study further applied the same concept to the factors corresponding to the four parameters rated from 1 to 10, resulting in two distinct profiles for each alternative and allowing for a more comprehensive analysis of the role of perceptions and experiences in mode choice. The sorting of colors associated with modes follows a roughly descending order



Figure 3.4: Median of the answers to 5-point Likert scale questions regarding experience, by mode choice.

of environmental friendliness to enhance this representation further. Finally, the study considered three measures of central tendency to represent the results for each profile: mean, median, and mode. From these measures, the visual assessment of differences between transportation alternatives is best achieved by the median for the 5-point Likert scale questions and the mean for the remaining ones. Although the median might hide some of the variability between the modes, it allows the most straightforward analysis of the first group of questions without comprising the slight differences between some modes. Conversely, the study proposes the mean approach for the remaining questions, as median and mode created, in such a case, highly complex overlapping patterns without displaying the differences between modes as efficiently as the chosen measure.

5-point Likert scale questions Figure 3.4 illustrates that, if excluding bicycle/scooter, which their users consistently perceive as the *ultimate traveling experience*, the modes differ primarily in whether individuals perceive them as their first choice of travel. Ratings on public transportation and shared modes potentially reflect their relative functionality, notwithstanding walking and own car trumping them as the first choice.

10-point Likert scale questions In its turn, Figure 3.5 provides evidence that public transportation, which is leaning environmentally friendly, is highly efficient, including in the economic dimension. Nonetheless, it arguably lacks the flexibility and sense of control most other alternatives provide. Active modes, typically deemed greener alternatives, appear more flexible and cheaper, contrary to own car, which offers more autonomy at higher monetary and ecological costs.

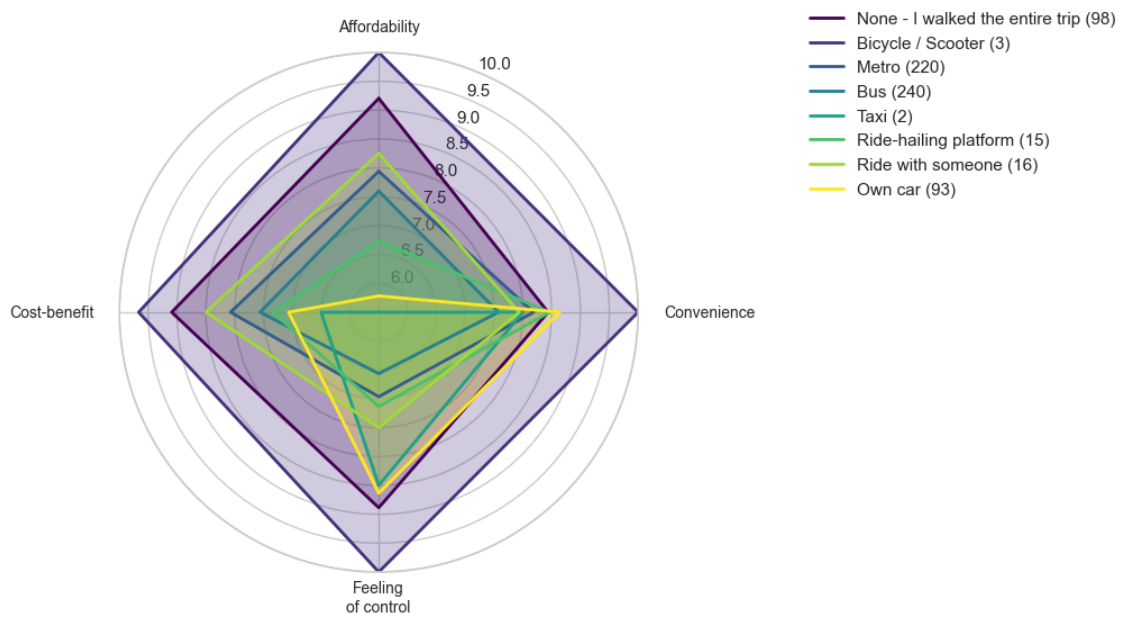


Figure 3.5: Mean of the answers to 10-point Likert scale questions regarding experience, by mode choice.

Likert scales and linear approximation The risks and limitations described by Wu and Leung (2017) of treating 5-point Likert scale responses as linear data (1 to 5) are recognized, yet practical constraints necessitated such approximation to facilitate discrete choice analysis. The exact approximation was used for the questions using 1 to 10 ratings, where the risk diminishes by employing a larger scale Wu and Leung (2017), but also since participants already provided their answers in numerical values.

Overall criteria for inclusion/exclusion As reported in Table 3.3, several pairs of these factors showed upper-moderate (e.g., satisfaction and enjoyment) to strong (e.g., satisfaction and happiness) positive correlations between themselves, suggesting high levels of redundancy. This analysis results in only keeping satisfaction and perception as first choice from the set of factors derived from the 5-point Likert scale questions. The first could work as a general proxy of trip experience, and the second was the factor that most differentiated each mode when visualizing the data.

Inside the remaining group of factors, perceived affordability moderately correlated with cost-benefit. Nevertheless, the first also correlates with trip cost, which is a significantly more objective and robust indicator of monetary expense related to the trip. Consequently, the study excluded both factors from further investigation. Convenience and feeling of control showed, although weak, some degree of correlation with choice and were included in the final set of factors, presented and further described in Section 3.3. This set was the one used for the initial parameter estimations.

3.2 Imputation method and validation heuristic

Given that a significant percentage of responses to several survey questions were *I don't know* / *I prefer not to answer*, such answers were treated as fields corresponding to missing data. There were also questions conditionally not applicable to some participants. Such a scenario resulted in 637 missing values, representing 3.57% of the total data. Of the 26 variables, 12 have missing values. The five variables—here corresponding to factors of choice—with the highest proportion of missing data are, in descending order: waiting time (30.71%), trip regularity (26.78%), walking time (16.74%), income level (10.19%) and cost (5.82%). These percentages are easily explained: for waiting/walking time, the questions were not applicable for all alternatives; for regularity, the question was also not applicable for participants who reported an occasional trip; for income level, there seems to be a natural reluctance in sharing such information, with more than 70% of missing values corresponding to *I prefer not to answer*; regarding cost, the missingness relates to a high proportion of *I don't know* answers, most present in car alternatives—in fact, this is the most chosen answer from car owners, at ~35%. There are 355 survey answers with at least one missing value, leaving 332 complete answers. These data are highly valuable, even if some fields are missing. Therefore, given their use for the further modeling phase, it was imperative to perform missing value imputation in the most accurate way to avoid dropping any data.

Not applicable responses Not applicable responses in the survey were represented as missing values rather than zeros. In reality, if using zero as a placeholder, trip regularity, waiting time, and walking time would possess only 3, 1, and 0 missing values, significantly reducing the total missing data to only 0.73%. Although this decision does not impact the parameter estimation directly, it may do so residually and indirectly since, with two of the proposed solutions for imputation, the imputed value for a given variable depends on values from other variables.

In practice, with such imputation techniques, what happened to missing values of variables like cost is that they were predicted using 14 fully complete and 11 incomplete variables instead of being predicted using 17 fully complete variables and 8 incomplete ones. This scenario comparison highlights the trade-off introduced by this decision. Finally, it is noteworthy that 5 out of the 11 variables to impute have only 1 missing value, followed by owning a subscription with 5 (0.73%), feeling of control with 7 (1.02%), and all the aforementioned ones after them. While this decision is arguably not optimal, the potential impact on the final results is likely minimal. Moreover, it is not as concerning as it could have been in a distinct context and a different missing data pattern. In addition, except for mode imputation—because it does not need it—all the approaches henceforth described ensured that missing values were adjusted to reflect only the existing values present in the original data by rounding, which inherently involves a certain level of approximation.

Across the most popular missing data imputation techniques, it was decided that a simple imputation method—as in a single value used as a placeholder for the whole missing values in a given variable—or, if needed, a more complex one—e.g., multiple or conditional imputation, which considers relationships and patterns among different variables—would suffice. Therefore,

this study proposes two alternatives for each type of technique. The present study employed Mode Imputation (MOI) and Median Imputation (MEI) for the more straightforward approach. Mean Imputation was likely inappropriate for the present case since the imputed values would not represent possible answers or translate gross approximations. Regarding the options with higher complexity, Multivariate Imputation by Chained Equations (MICE) and Random Forest Imputation (RFI) are prominent techniques to tackle the problem. Other options like K-Nearest Neighbors Imputation were considered, though dropped after a series of implementation issues—probably due to not being the most well-suited algorithm for only partially numerical data.

Mode and median imputations On the one hand, with the first alternative, missing values were replaced with the most frequent value—or the most common category, in the case of categorical data—inside the respective variable. On the other hand, before value mapping, i.e., assigning the actual label or numeric value to each element, each variable with missing data was sorted to select its central value. Note that, for MEI, the applicability to nominal data—corresponding to factors trip character and sex—is limited, as there is no inherently meaningful order between their respective categories. Despite that, applying this imputation across the whole data was still appropriate, given that the mentioned factors contained only 1 missing value each, and differences in imputation results were expectedly unnoticeable for the end result. Furthermore, imputation would always be somewhat ambiguous since only two categories were available for both variables. Datasets MOI and MEI resulted from these techniques.

Multivariate imputation by chained equations MICE is a robust missing data imputation algorithm that operates iteratively, working on filling in one variable at a time using the remaining ones. It works according to van Buuren and Groothuis-Oudshoorn (2011) and, although initially created for R programming language, it is alternatively available for Python as part of the scikit-learn library, under the `IterativeImputer()` function. The present study preferred this alternative, implemented with all the function's arguments set to default.

The initialization step consists of imputing all the missing data at random. However, using an initial approximation strategy like mean, median, mode, or constant imputation is also typical. Here, the default is mean imputation. Then, each variable's approximated values are substituted by predicting its actual value with a conditional algorithm most suited to the characteristics of the data. Again, the default prevailed, corresponding to Bayesian ridge regression. After all variables are processed, an iteration ends, and the cycle continues until a predefined number of iterations is met. No more than five iterations are usually necessary. Therefore, the specified parameter for the algorithms' maximum number of iterations was 10—which is also the default for the used function. The whole process runs multiple times, producing a set of distinct complete datasets to account for stochasticity. The function used, however, has a fixed deterministic design to result in a single solution, therefore providing only one dataset with highly plausible imputed values.

Missing data mechanism There are different mechanisms of missing data, as classified by Rubin (1976). It is important to note that MICE assumes the data follow one of these mechanisms, i.e., is Missing At Random (MAR). This assumption holds when the probability of a value being missing in one variable is independent of missingness in other variables but may depend on the observed values of the variables in question. If this condition applies, the missing data mechanism can be considered *ignorable*, meaning that the reason for the missing data is unimportant; only MAR occurs. For instance, younger people may feel uncomfortable sharing their income level. MAR would be satisfied if the likelihood of missing values in income level were related to age but unrelated to income level within a given age interval. In this proposed example, within the group of younger people, whether or not someone's income is missing is unrelated to how high or low their income is, as long as data are MAR (Nisbet et al., 2018a).

The presence of these mechanisms can be determined by visualizing missing data patterns. Suppose one observes a plot like the one in Figure A.19 of Appendix A and red cells are randomly scattered and sparse. In that case, it suggests data are Missing Completely at Random (MCAR), i.e., the probability of a value being missing in a given variable is entirely unrelated to observed and unobserved values. If red cells concentrate in specific patterns—e.g., grouped by variable or missing in combination—it points to MAR. Finally, if the missing values in a variable consistently appear in a specific subset, it could indicate a third mechanism: Missing Not at Random (MNAR), where missingness is related to unobserved data. This last scenario is significant for MICE because the algorithm cannot capture unobserved dependencies and thus becomes unreliable if data are MNAR. In the present case, as shown in Figure A.19 of Appendix A, a mixture of structured blocks of missingness and independently missing blocks suggests a potential MAR mechanism. Solely based on visual investigation and without further examination, data can be assumed as MAR for practical reasons. Accordingly, this study implemented the MICE algorithm, resulting in the homonymous dataset.

Random forest imputation RFI, in contrast, is non-iterative, meaning it is performed once for each variable of the whole dataset, independently—not conditionally—from other variables' imputed values. This feature implies that the imputed variables do not update based on the imputation of another variable once predicted for the first time, resulting in a more straightforward and less computationally expensive process.

First, the data splits into *observed* and *missing*, respectively, corresponding to the rows where a target variable to impute has data and where the data is missing. Likewise, the procedure treats the remaining variables as the predictors. Considering these definitions, a Random Forest (RF) model uses the *observed* data for training. After this stage, by fitting the model to the predictors' observations present in the *missing* set, the procedure can fill in the missing values of the target variable. Since this method assumes that only one variable contains missing values each time, a simple mode imputation (or equivalent) must be applied to the predictors before running the imputation, making it even less robust than MICE. When all the missing values are imputed (via

Table 3.5: Factors included in the modeling phase.

Factor
<i>Socio-demographic</i>
Age
Sex
Household composition
Owning a subscription
Income level
<i>Objective</i>
Is the trip regular?
Trip regularity
Purpose
Number of modes
Trip duration
Cost
Walking time
Waiting time
<i>Subjective</i>
Satisfaction
Convenience
Feeling of control
Perception as first choice

This set of factors is the same as the one outlined in Table 3.9 under the name *Full*.

model fitting), the resulting dataset combines those values with all the observed data, resulting in a complete dataset.

This study eventually dropped the described approach to overcome its limitations while ensuring fair comparisons inside these more complex options. The procedure was changed to reflect what is, in fact, a MICE implementation with RF as an estimation procedure instead of the previous Bayesian ridge regression. Henceforth, the present study denotes the dataset generated by such a procedure as the RFI dataset.

3.2.1 Validation rationale

This study proposes a heuristic for model validation. Its chief purpose and concern are to measure the different model configurations experimented with, as depicted in the following sections of this chapter. Nonetheless, before describing the rationale behind such a procedure, the present section introduces a few notions on the types of validation contemplated.

Likewise, the heuristic must also serve validation in the context of missing data imputation. Hence, the four generated datasets with the different approaches described (MOI, MEI, MICE, and RFI datasets) underwent such a validation to fully assess the best imputation technique for the data. To do so in the context they will be later used, the most simple and standard DCM (MNL) was selected. Since this model is used in this section instrumentally, the study refers to Section 3.3 if a more precise specification of the said model is required. All the datasets were validated using

a baseline set of factors to ensure fair comparison; this set comprises the factors included in the modeling phase as described previously. Table 3.5 summarizes its composition.

In-sample validation vs. out-of-sample validation In-Sample Validation (ISV) typically trains the model in the entire dataset to determine its learning capabilities regarding the data patterns, indicating how well it fits such data. Since it does not measure generalizability for unseen data, these initial indications of the model's performance often overestimate actual performance. Overfitting can also be a problem if this is the sole validation reported, as the tendency is, via higher complexity, to capture more and more of the noise and irrelevant details of the data and not the general, most important, patterns.

Out-of-Sample Validation (OSV) differs from ISV since it evaluates model performance in a separate dataset, the validation set, containing data unused in the training process—the so-called unseen data. The most significant advantage of this technique is its agility, which provides a more realistic assessment of the model's quality and proneness to overfit/underfit. Holdout Validation (HOV) and Cross-validation (CV) are types of OSV.

Holdout validation vs. cross-validation A HOV approach evaluates a model's performance on unseen data and is best known for its application to machine learning implementations. This process splits the dataset into training, validation, and/or test sets. The model is trained on the training set and validated/tested on the validation/test set. It is simple and computationally efficient, although slightly more prone to overfitting due to the reliance on a single data split.

When the HOV process is repeated multiple times with randomly generated training-validation set pairs, the validation procedure refers to the Repeated Holdout Validation (RHOV) approach. Suppose this split is done systematically into v folds, ensuring every data point is used exactly once for validation/testing and $v - 1$ times for training. In that case, the method is called cross-validation (CV). This more robust technique reduces the variability and bias associated with the single split. Nevertheless, it can be overly complex in computation for larger datasets and/or heavier algorithms (Burman, 1989).

External validation vs. internal validation External validation proposed by Parady et al. (2021) was ultimately impossible since the working data were not one of the following: from different periods, distinct regions, and gathered with various methods. Even so, many of the model choice factors presented are context-specific, as is how the survey puts its questions to extract the information. Data variability is another complex issue to get around, e.g., expressed in different demographic patterns or factor scales, making the model automatically less effective. Such is also extremely important because the theoretical assumptions used to choose the best factors for a given model might not hold in a different context: a factor deemed likely irrelevant before parameter estimation with the present data has the chance of being exceptionally relevant when working with data with the same—at least, semantically—factors. Therefore, the transferability of the estimated

models could not be assessed—contrary to their reproducibility. With that in mind, the generalizability of these models was exclusively based on the extent to which they could sustain predictive power in different samples from the same population.

The limitations inherent to this fact must be acknowledged—such as bias and risk of overfitting—yet it should not be understated the fact that, solely by reporting validation performance measures, the ability to infer on policy measures is significantly enhanced, as this represents a far more robust approach than simply reporting goodness-of-fit for a model estimated on every single data observation. Finally, it is worth mentioning that the present study fits inside a small percentage (18.1%) of studies that opted for this approach (Parady et al., 2021), allowing for higher confidence in the research’s credibility.

The proposed validation heuristic Partially following the recommendations by Parady et al. (2021), the validation of each estimated model was divided into a two-fold approach combining ISV and OSV, with the latter corresponding to internal validation using HOV—data split into 80% used for estimation and 20% used for validation—in detriment of a CV procedure, which would considerably increase computational demands, particularly for nested structures or LCMs. Although opting for this ratio was a slightly arbitrary decision based on typical practices outside discrete choice modeling, according to the summary of articles reporting validation in the literature listed by Parady et al. (2021), the majority of studies that report internal validation measures opt for HOV (18 out of 32) and the most common split ratio is precisely 80-20 (8 out of these 18).

Note that the chosen ratio of 80-20 implies the non-existence of a dedicated testing phase, as dividing the data into three different portions would significantly reduce the data available for estimation (training) even more. Furthermore, final testing with rigorous detail regarding individual classification is not as urgent as interpreting the estimated coefficients for the whole sample or, for instance, evaluating the aggregate outcomes of the model on a portion of such a sample. For the latter, the 20% dedicated to validation must suffice, even if at the cost of some minor level of overfitting.

3.2.2 Validation metrics

This study presents two distinct sets of metrics for the two validation steps. Nonetheless, both steps share some of them, particularly probabilistic-based metrics related to the model’s fit and predictive accuracy. Although computed the same way in both validation steps, their level of importance inside each step might differ. ISV mainly focuses on balancing goodness-of-fit and model complexity while assessing its apparent predictive accuracy and the confidence of predictions. Conversely, while the previous is primarily applicable for OSV, the focus shifts to the model’s ability to maintain classification performance in a different sample and estimate the respective market shares on the said sample, i.e., the validation sample.

Metrics for in-sample validation only

For ISV, it is more important to gauge goodness-of-fit and the model's overall quality regarding the balance of fit and complexity since the apparent predictive accuracy is, as the name implies, simply apparent—nonetheless, still important.

Model selection criteria In terms of model selection criteria, the Akaike Information Criterion (AIC) (Akaike, 1974) and Bayesian Information Criterion (BIC) (Schwarz, 1978) were computed according to their modern formulations, as described in Burnham and Anderson (2004).

Prediction confidence Percentage of Clearly Right (%CR), Clearly Wrong (%CW), and Unclear (%U) predictions, which indicate the level of confidence in the model's predictions, were calculated in line with the definitions proposed by Parady et al. (2021) and the probability threshold used is 0.75. These last three measures could have fit OSV as well. However, this study does not report them in such a step for two reasons. First, what they measure can be extrapolated as the same or, at least, similar for the model estimated on OSV's training data since both belong to the same *system of models* (the discussion of the metrics only applicable to OSV addresses the meaning behind it). Second, in OSV, the aggregate outcomes of the model are already known, and in the case of a high %CW, the overestimation/underestimation of the market shares for some alternatives—particularly for those with the highest observed shares—will likely already reflect the said high %CW. In short, the focus narrows toward other statistics.

Metrics for both validation steps

As indicated earlier, not all metrics are exclusive of each validation step. For instance, every time a parameter estimation is performed, the Log-Likelihood (LL) of the estimated model and the number of estimated parameters must be reported. K represents this number of parameters the model estimates.

Log-likelihood-related metrics The LL measures the goodness-of-fit of a model to a given dataset. As its name implies, it derives from the likelihood function, i.e., the function that expresses the probability of the observed data given the model's parameters. The computation of this measure is further described in Section 3.3 when discussing details of the model's estimation procedure. On ISV, $LL_e(\hat{\beta}^e)$ is the LL of the model estimated on all observations, and $LL_e(0)$ is the log-likelihood of the null model (NLL) for the same data, i.e., the log-likelihood of the model when all parameters are zero. On OSV, $LL_t(\hat{\beta}^t)$ is the LL of the model estimated on the training data, and $LL_v(\hat{\beta}^t)$ the LL of the model estimated on the training data applied to the validation data; $LL_t(0)$ is the NLL associated with the training data, and $LL_v(0)$ the NLL associated with the validation data.

The likelihood ratio index (ρ^2) (McFadden, 1974), also known as McFadden's R^2 , was computed according to Train (2009d). The adjusted version of the same statistic, which introduces a

penalization for the number of parameters in the model, known as Adjusted McFadden's R^2 ($\bar{\rho}^2$) (Ben-Akiva and Lerman, 1985), was computed as described in Long (1997). Higher ρ^2 and $\bar{\rho}^2$ values are typically better, with those closer to 1 corresponding to the state where the estimated parameters nearly perfectly predict the choices of the sampled decision-makers (Train, 2009d).

The study also employed the likelihood ratio test (LRT) when applicable, i.e., when comparing *nested* models, as exemplified by Train (2009d). The reader should note that the term *nested* is unrelated to the previously presented extension of the DCM: one model is *nested* within another model when the first (known as *restricted*) is a special case of the second (known as *unrestricted*). If such is not the case, an alternative to this comparison is performed through the Vuong test (VT) (Vuong, 1989), in addition to evaluating AIC and BIC. For every test mentioned, 5% was the statistical significance level used.

Accuracy The percentage of correct predictions, also known as first preference recovery, denoted in this study as accuracy (ACC), was computed, again, in line with the definitions proposed by Parady et al. (2021), using a deterministic approach: the study assumes the predicted choice is the alternative with the highest predicted probability.

Metrics for out-of-sample validation only

Secondarily, for OSV, it is crucial to analyze how a *system of models* maintains its classification performance in different samples and how well it fits the training data and the data present in a new sample, i.e., the validation data. This reference to *system of models* in OSV is essential to have in mind because models estimated with the same specification and the same set of factors will differ when the percentage of data used for training is different. A model estimated with 100% of data (as in ISV) will most likely never result in the same parameter estimation—they do not represent the same maximum likelihood estimate—as a model estimated with 80% of the analogous data (as in OSV). Hence, it could be said that they belong to the same *system of models*, for they share identical characteristics, yet they are not the exact equivalent model.

Classification Precision (PREC) is the proportion of correct positive predictions, and F1-score (F1) is the harmonic mean of PREC and recall (TPR), as presented by Nisbet et al. (2018b). Since this is a multinomial context with an imbalanced choice set, each class's varying support (number of instances) must be considered. Consequently, the reported values represent the weighted average of each metric. If not, the results could represent overestimations of actual performance. For example, the ACC registered in ISV of LCM estimated on BS2 (see Table 3.9) would be 0.92 instead of the reported 0.88 (see Table 3.13). A noteworthy remark is that TPR is technically reported under ACC because both reduce to the same mathematical formulation in the context of weighted averages. All these metrics use the predictions made under the premise used to compute ACC.

Market shares The study utilized these same predictions to compute each alternative's market shares, i.e., the ratio of each alternative in the validation data. This sample is the same for every validation report and contains 138 observations with the observed market shares shown in, e.g., Table 3.8. Moreover, the average of the absolute errors between predicted and observed market shares was the mean absolute error (MAE) formulation.

Mean absolute error vs. mean absolute percentage error MAE was preferred to mean absolute percentage error (MAPE) because a globally minimal market share difference would result in overestimating such a metric. If, e.g., the model fails to predict all the three individuals observed in the validation sample who chose ride-hailing, ride-hailing's absolute percentage error (APE) enters the computation of MAPE under the value of 100%, leading to the mentioned overestimation. MAPE would have to be computed as the weighted average of APEs to weigh the class imbalance. Such culminates in none other than the formulation of MAE. Furthermore, it would not be possible to compute the APE of taxi since the market share of this alternative in the validation sample is 0%. Additionally, the way each model predicts each alternative individually is always analyzed ("Model X is the best at predicting the market share for alternative Y"). If needed, the reader could inspect Parady et al. (2021) for additional details on types of errors typically reported in the literature.

Deterministic vs. probabilistic approach for prediction

The criticism of using ACC as a validation metric or goodness-of-fit statistic, as presented by Train (2009d), was most certainly present at the time of implementation. The author states that the way to compute the metric is opposed to the meaning of probabilities since its basis is a deterministic premise. As an illustration, 100 out of 100 times, the individual would make that same choice, which is not factual. The consequences of such a premise might also affect the computation of market shares.

While this determinist approach implies such an assumption, the simplification is practical and offers otherwise impossible advantages at a disaggregate level. In fact, by complementing ACC with PREC and F1, the model's classification performance can be assessed at the individual level. This assessment is only feasible if the deterministic assumption is made *a priori*. Suppose the model consistently fails to predict any alternative with a low average choice probability. In that case, such an issue is always observable in the prediction success table, and the analysis or model adjustments are made accordingly. Besides, using the same approach to measure aggregate and disaggregate outcomes allows for higher coherence of analysis since ACC, PREC, and F1 can reflect some of the patterns observed in market shares but also unobserved. For instance, the model might predict that one individual chose bus when they chose metro while predicting that another individual chose metro when they chose bus. Even though both predictions are incorrect, their opposing errors balance out, reflecting a correct market share forecast. Although such is not as important as measuring the overall travel demand for each alternative, it is certainly advantageous to have in mind.

Table 3.6: In-sample validation results from the datasets generated with each imputation technique.

Dataset	ρ^2	$\bar{\rho}^2$	$LL_e(\hat{\beta}^e)$	AIC	BIC	ACC	%CR	%CW	%U
MOI	0.6180	0.2701	-208.6381	797.2762	1658.420	0.8632	68.12	4.51	27.37
MEI	0.6175	0.2696	-208.9254	797.8508	1658.994	0.8617	67.83	4.51	27.66
MICE	0.6255	0.2776	-204.5436	789.0872	1650.231	0.8632	67.54	3.93	28.53

These results were obtained with the MNL as described in Section 3.3 using all possible factors (see Table 3.5).

For all datasets:

$K = 190$

$LL_e(0) = -546.1821$

Furthermore, the validation approach does not use statistics alone. Instead, probabilistic metrics like LL, ρ^2 , or $\bar{\rho}^2$, which explicitly capture the probabilistic nature of the model, always complement them, as described previously. Additionally, the predictions—made under each opposing approach—tend to converge to the same value as the sample size increases as long as the estimated probabilities show enough randomness. The present sample where these market shares are evaluated is arguably big enough ($N = 138$) for the differences between them to become non-substantial.

3.2.3 Validation results

The validation results on the examination of the best-imputed dataset appear in Tables 3.6, 3.7, and 3.8.

In-sample validation

The MICE dataset (dataset with missing data imputed through MICE) reported the best ISV results in almost every metric, except for %CR and %U, as demonstrated by Table 3.6. Specifically, regarding goodness-of-fit, it performed well. Since all models used the same set of factors, the LL can be directly compared without resorting to LRT or VT (also applicable to OSV), with MICE presenting the lowest value (-204.5436). From this fact, it is intuitive to derive that it also presents the best ρ^2 and $\bar{\rho}^2$. According to ρ^2 (0.6255), the model can explain 62.55% of the improvement over the null model—where all the alternatives' likelihood is the same. According to $\bar{\rho}^2$, however, with the penalization for the number of parameters of the model, the value drops to 0.2776, suggesting a reduced fit efficiency via high complexity. Although indicating considerable room for improvement, it remains an acceptable figure.

Its predictive accuracy is also satisfying, with an ACC of 86.32%, representing almost 7 times the baseline accuracy for random choice in this context when all modes are available since there are eight alternatives (1 divided by 8 equals 12.5%). For the %CR, %CW, and %U metrics, the set probability threshold was 0.75, which is, henceforth, the same for every model reported to ensure consistency. This study decided to use this threshold based on a reasonable premise: it portrays 3

Table 3.7: Out-of-sample validation results from the datasets generated with each imputation technique.

Dataset	ρ^2	$\bar{\rho}^2$	$LL_t(\hat{\beta}^t)$	$LL_v(\hat{\beta}^t)$	ACC	PREC	F1
Training Data							
MOI	0.6582	0.2330	-152.7470	-	-	-	-
MEI	0.6577	0.2325	-152.9655	-	-	-	-
MICE	0.6746	0.2494	-145.4096	-	-	-	-
Validation Data							
MOI	-1.7156	-3.6285	-	-269.7209	0.8116	0.8184	0.8120
MEI	-1.5326	-3.4456	-	-251.5498	0.8116	0.8184	0.8120
MICE	-2.6772	-4.5901	-	-365.2298	0.8043	0.7953	0.7989

These results were obtained with the MNL as described in Section 3.3 using all possible factors (see Table 3.5).

For all datasets:

$$K = 190$$

$$LL_t(0) = -446.8581$$

$$LL_v(0) = -99.3240$$

times the probability of random choice for an individual with four available alternatives (half the eight total, in this case) in a context of equal market shares. Here, a %CR of 67.54% for MICE represents more than 5 times the baseline accuracy for random choice, although smaller than the 68.12% resultant from MOI. These percentages mean that the models correctly predict such percentages with high confidence. To compensate, MICE presents less clearly wrong predictions (3.93% vs. 4.51%).

Out-of-sample validation

The situation slightly changes when looking at OSV, as shown in Table 3.7. The goodness-of-fit for the MICE training data is still the highest, although the same does not apply to the validation data. Since the MLE's optimization algorithm failed to converge when using the training data

Table 3.8: Validation sample's market shares (%) from the datasets generated with each imputation technique.

Dataset	Bus	Metro	Own car	Walking	Ride-hailing	Ride	Bicycle/Scooter	Taxi	MAE
Market Shares - Observed									
-	34.06	33.33	14.49	12.32	2.17	2.17	1.45	0.00	-
Market Shares - Predicted									
MOI	30.43	36.96	14.49	12.32	2.90	2.17	0.72	0.00	1.09
MEI	30.43	36.96	14.49	12.32	2.90	2.17	0.72	0.00	1.09
MICE	31.88	35.51	14.49	13.04	2.90	2.17	0.00	0.00	0.91

These results were obtained with the MNL as described in Section 3.3 using all possible factors (see Table 3.5).

Ride represents *riding with someone*.

from the RFI dataset, this study does not report any validation results from the said dataset. The reasons for this lack of convergence are analyzed and discussed in Section 3.5.

The problem with ρ^2 on the validation data These values are consistently negative throughout almost all validation reports, pointing out that the models may possess poor reproducibility and might be overfitting to the training data. This fact does not immediately invalidate the models. Instead, the statistical translation is that these are simply models that perform worse than the null model in explaining the variance of the validation data ($LL(\hat{\beta}) < LL(0)$).

As pointed out by Train (2009d), theoretically, the lowest value ρ^2 can take is 0 and, coherently, $LL(\hat{\beta}) < LL(0)$ is not possible because it would correspond to a model with coefficients that do not represent the maximum likelihood estimate for the observations. This scenario happens in reality and explains the negative values: the coefficients fitted to the validation data were estimated on the training data; hence, they do not represent the MLE for the validation data. In this specific validation context, reading positive values for the measure would be preferable, but the over-reliance on goodness-of-fit must stop at some point. Put into context, the result is not problematic: in this step, the reported classification measures and the model's discriminative ability in aggregate terms are also essential aspects to consider. As long as these measures do not degrade significantly from ISV to OSV, there are no considerable limitations. Likewise, as mentioned, this is not without exception, as some estimations presented afterward show the limitation settled.

Ultimately, the dispute is between the MOI and MICE: while the first performs better regarding fit to validation data, the second does so for the training data. In contrast, the first presents slightly better ACC, PREC, and F1; the predictions made by the second align better with observed market shares. It presents the lowest MAE (00.91%) and is the best predictor of the shares for the most significant modes (bus and metro), to the detriment of bicycle/scooter and walking.

Overall analysis

The MICE dataset achieved superior goodness-of-fit, model parsimony, and prediction performance balance, including better alignment with the observed market shares. Despite negative ρ^2 values for the validation data, which will be further addressed, the *system of models* estimated on the MICE dataset demonstrates that such is the most robust dataset with a high degree of certainty. There is potential for further model improvements, as shown in Section 3.3. The MICE dataset was used in the subsequent stages of the modeling phase to proceed accordingly. Consequently, the present study excluded the remaining datasets (MOI, MEI, and RFI) from further analysis.

3.3 Multinomial logit model

As discussed previously, an MNL is grounded in the random utility theory, which expects decision-makers to choose an alternative providing the highest utility for themselves—this is the so-called *homo economicus*, a perfect optimizer characterized by a tremendous ability to make rational, economic-inclined, decisions. The first DCM implemented builds on the basic specification of the

random utility model. Using Eq. 3.1 as a reference, the utility function can be rewritten explicitly according to the following format:

$$U_{in} = V_{in} + \varepsilon_{in} \quad (3.2)$$

$$V_{in} = \beta_0 i + \beta_1 \cdot x_{1,in} + \beta_2 \cdot x_{2,in} + \cdots + \beta_k \cdot x_{k,in} \quad (3.3)$$

As shown in Eq. 3.2, total utility U_{in} is the sum of a *deterministic part of utility* V_{in} (also known as *systematic component of utility* or simply *systematic utility*), based on measurable factors, and a *random term* ε_{in} (also known as *random variable* or *error term*), based on the unobservable factors—nothing more than what the model cannot explicitly account for. In this sense, V_{in} can be modeled as a function of attributes of alternative i and individual-specific characteristics for each individual n , usually socio-economic and/or socio-demographic. Such is presented in Eq. 3.3, with these attributes and characteristics merged into a single group of attributes x_{in} (the so-called *factors of choice*).

The alternative specific constant For each mode alternative, β_0 represents the mean of the random variable ε_{in} . The study further presents it as *ASC*, which stands for *Alternative Specific Constant* (ASC), i.e., a specific parameter that captures the systematic inherent bias for alternative i not explained by the K factors contemplated by the model. Therefore, ASC acts as a placeholder for factors not explicitly included, but that systematically affect the alternative's utility. For instance, if own car is consistently preferred, ASC_{OwnCar} reflects this systematic bias. Eq. 3.3 can be rewritten as:

$$V_{in} = ASC_i + \beta_1 \cdot x_{1,in} + \beta_2 \cdot x_{2,in} + \cdots + \beta_k \cdot x_{k,in} \quad (3.4)$$

Although the ASC is conceptually part of random utility, it can move to the deterministic component. When included in the utility specification, the random term ε_{in} can be normalized to have a mean of zero without loss of generality. This normalization is a standard assumption of DCMs, meaning that, on average, ε_{in} does not systematically bias total utility U_{in} towards any specific value. Under such assumption, V_{in} represents the mean of the total utility U_{in} , thus being seen as the *expected utility* and including the mean of any systematic biases as, e.g., the ASC. Because only differences between utilities drive choice probabilities, differences in means—as reflected in V_{in} —are what truly matter.

Therefore, everything else being equal, this parameter, given its positive/negative sign and magnitude, informs the research about the choice the individual n is more likely to perform via higher baseline utility. In the present context, the study arbitrarily normalized the ASC for the bus alternative (ASC_{Bus}) to 0, therefore representing the reference alternative. The logical consideration was that it is the most common choice inside the working sample.

Firstly, this is achievable because—once more—only the differences in utility matter to determine the probabilities related to each alternative: absolute utility values have no inherent meaning, given that probabilities depend on the relative attractiveness of one alternative compared to others.

Secondly, doing so aids in avoiding redundancy in parameter estimation and ensures the model is not over-parameterized. Thirdly, normalizing the ASC related to bus to 0, e.g., is the same as saying that such alternative represents the *status quo*, or even better, a *neutral choice*, facilitating the straightforward interpretation of results: a positive ASC for metro would mean metro is preferred relative to bus, *ceteris paribus*.

The error term and probability formulation Unobservable attributes and measurement or specification errors are typical sources of the error term ε_{in} . The literature typically presents two distinct families of models based on their respective assumptions about ε_{in} . The *Probit* family assumes that ε_{in} follows a normal distribution, which results from the aggregation of many independent sources of error (consistent with the Central Limit Theorem). The *Generalized Extreme Value* (GEV) family, which includes the logit model, assumes that ε_{in} follows an Extreme Value distribution (EV) (Gumbel, 1958), which results from assuming that ε_{in} captures the largest of many independent sources of error. In the logit framework, ε_{in} for all alternatives i is also assumed to be independent and identically distributed (IID).

When evaluating the difference of the errors associated with two different alternatives for the same individual (ε_n), the problem is finding the distribution of such difference. Nonetheless, from the properties of EV, it can be derived that ε_n follows a logistic distribution. This distribution's cumulative function underpins the probability expression of the logit model. Using the previous assumptions and the properties of the logistic distribution allows the probability of individual n choosing alternative i under the MNL to be generalized and simplified to the following formulation:

$$P_{in} = \frac{e^{V_{in}}}{\sum_{j \in C_n} e^{V_{jn}}} \quad (3.5)$$

Not all alternatives may be available or relevant for every individual n . Since the implemented model takes into account the availability of each alternative i for each individual n , the probability computation uses the choice set C_n of the only available alternatives for that same individual.

Independence of irrelevant alternatives The IID assumption of error terms leads to the independence of irrelevant alternatives (IIA) property, which is well-known as a typical limitation of MNL. This key property of MNL defines the relative odds of choosing between two alternatives as constant, regardless of the presence/absence of other alternatives in the choice set. In other words, adding or removing a third alternative to the choice set does not affect the ratio between the probabilities of the other two alternatives. However, this may not always be realistic, for some alternatives inherently compete with each other at a higher degree, e.g., bus and metro might be closer substitutes than bus and own car. This real-world scenario often violates the IIA assumption, which is only relaxed by changing the model's configuration—e.g., using an NL or LCM.

Final utility specification In the first specification described, the parameters β_k , corresponding to the coefficients of factors k , were the same for each alternative. In practical terms, this ultimately means that the model will assume each factor affects each alternative similarly. Such behavioral assumptions can be dangerous, as a factor like cost is likely to affect bus users and car owners differently. In fact, after initial estimations—which this study does not report—the yielded results made it clear that such was the case, as the models estimated extremely small coefficients for each factor, making them irrelevant for the utility computation across all alternatives. Such means that what would ultimately differentiate the different modes was their respective ASCs. Furthermore, a distinct specification where the coefficients related to individual-specific characteristics were the same for every alternative was experimented with (since these characteristics do not vary over alternatives). Still, the same problem arose specifically for these coefficients.

Logit models can capture taste variation within certain limits. In particular, tastes that vary systematically according to observed factors can be incorporated into these models. However, tastes that vary with unobserved factors or purely randomly cannot be handled. All in all, the study relaxed the initial assumption for all factors k among all alternatives i , and such was the format of the final utility specification for the deterministic part:

$$V_{in} = ASC_i + \beta_{1,i} \cdot x_{1,in} + \beta_{2,i} \cdot x_{2,in} + \cdots + \beta_{k,i} \cdot x_{k,in} \quad (3.6)$$

In the new specification, the attribute's coefficients are segmented by alternative. Thus, the model postulates that, in behavioral terms, the same individual n perceives any factor k differently for each alternative i . Hence, for each alternative equation, the estimated weight and sign of that same factor is not constant. This assumption is the same as saying the coefficients for each factor are, in similarity to ASC, *alternative-specific coefficients*. Although such cannot eliminate the IIA property—as it is a structural property of MNL—it may mitigate practical consequences by capturing preference differences across alternatives most adequately, which would otherwise be missing.

Maximum likelihood estimation After the specification of a DCM, the primary task is to estimate the coefficients of the unknown parameters. Maximum Likelihood Estimation (MLE) is the standard procedure for performing such a task. This statistical method finds the values of the parameters that maximize the likelihood function—which measures how well the model explains the observed data. Comprehensive descriptions of this method are consistently found in the literature (Bierlaire, 2003; Myung, 2003; Cosslett, 1981).

The key typical steps of MLE are as follows. First, *model specification* is required. This step involves defining the probability distribution (e.g., normal, binomial) that describes the data. The model depends on parameters β_k , which must be estimated. Hence, in the following step, the *likelihood function* $L(\beta_k|y)$ expresses the probability of observing the data y given the parameters β_k . Eq. 3.7 shows its definition. For independent observations $y_1, y_2, \dots, y_n$, i.e., the data points of

the studied sample, this $L(\beta_k|y)$ is computed according to Eq. 3.8.

$$L(\beta_k|y) = P(y|\beta_k) \quad (3.7)$$

$$L(\beta_k|y) = \prod_{i=1}^n P(y_i|\beta_k) \quad (3.8)$$

Nonetheless, to simplify computations and ensure numerical stability, the *log-likelihood function*, which is the logarithm of the likelihood, is preferred (see Eq. 3.9). The final step is *optimization*, where the procedure finds the parameter values $\hat{\beta}_k$ that maximize the log-likelihood function (see Eq. 3.10). This final step involves solving Eq. 3.11, which provides a necessary condition for maximization, and verifying sufficiency using the second derivative condition, as shown in Eq. 3.12. Typically, numerical optimization algorithms are employed to obtain $\hat{\beta}_k$, especially when closed-form solutions are not feasible.

$$\ln L(\beta_k|y) = \sum_{i=1}^n \ln P(y_i|\beta_k) \quad (3.9)$$

$$\hat{\beta}_k = \arg \max_{\beta_k} \ln L(\beta_k|y) \quad (3.10)$$

$$\frac{\partial \ln L(\beta_k|y)}{\partial \beta_k} = 0 \quad (3.11)$$

$$\frac{\partial^2 \ln L(\beta_k|y)}{\partial \beta_k^2} < 0 \quad (3.12)$$

In terms of estimation software, the implemented option was the Biogeme 3.2.14. package (also known as Pandas Biogeme), developed by Michel Bierlaire. (Bierlaire, 2003, 2023). This tool is an open-source Python package designed for the MLE of parametric models. For each estimation, the software generates a complete report gathering information about its results, including several of the metrics mentioned above used for model validation, among others (number of estimated parameters K , sample size N , NLL, LL, LRT, ρ^2 , $\bar{\rho}^2$, AIC and BIC). Moreover, the present study executed the estimation function provided by the mentioned package with the default parameters. It has *Newton with trust region for simple bound constraints* (as denoted in the report) as its default optimization algorithm. If needed, the documentation of the Biogeme package provides more details on the said algorithm. This study employed this software to estimate all previously described models except mixed logit (MMNL), which, in favor of the more straightforward latent class approach, was not included in the present study.

3.3.1 Definition of the sets of factors

Regarding the data and which set of factors to use, the study performed the first estimations using the factors said to be included in the modeling phase, i.e., all the factors outlined in Table 3.5 of the previous section. Although it contains 17 factors—which might already be considered excessive—a complexity issue arises since trip purpose and income level have 8 and 3 unique categories, respectively. In these cases, a base category was arbitrarily selected: *Study* for trip purpose and *No Income* for income level—categories with the highest frequency for the respective factors. This base category is normalized as a reference and excluded from the utility specification. The remaining categories are included as binary variables, resulting in a first set of 26 variables. The illustrated over-parametrization emphasizes the existence of the complexity issue mentioned.

In fact, at a statistical significance level of 5%, coefficients from 4 different variables showed no statistical significance for every single alternative. The variables in question were the convenience factor, plus some of the binary variables related to trip purpose, namely *family-related business*, *take children to school*, and *work*. Such potentially means that these variables might not influence the decision-making process and are irrelevant for the individual agents. However, their impact on the model's results arguably still exists in significantly residual terms.

Factor selection strategies

This study further implemented three different strategies of factor removal from the model, evaluating whether their omission would benefit computation, parameter estimation, and validation results: two significance-driven strategies, consisting of a simple backward selection (1) and backward selection with a tighter threshold (2), and a hypotheses-driven strategy built on the expected behavior in the present context (3). Table 3.9 presents the resulting sets of factors from all the employed strategies, which are explained as follows.

Simple backward selection The first strategy involved reestimating the model and analyzing the parameter estimation results from the specified MNL without interference from such factors. Here, concerning *trip purpose*, the followed approach was a cautious one, as removing only the non-significant binary variables could potentially affect the behavioral foundation of the model. In practical terms, it was also important to know the cost of the decision in terms of interpretation and, since the model still contemplated two mode-specific factors—*trip character* and *trip regularity*—, these would have to suffice as proxies for the whole sub-group of factors. Therefore, *trip purpose* was removed entirely, as was *convenience*, culminating in the *Backward Selection step 1* set of factors (BS1), as shown by Table 3.9. The process was repeated, and, in this second iteration, *household composition* and *sex* were removed. The elimination stopped after reestimating the model since no more factors could be removed without losing fit or explanatory power. The resulting set is denominated *Backward Selection step 2* (BS2), following Table 3.9.

Table 3.9: Sets of factors considered for parameter estimation.

Factor	Full	BS1	BS2	TT1	TT2	HD
<i>Socio-demographic</i>						
Age	Y	Y	Y	N	N	Y
Sex	Y	Y	N	N	N	N
Household composition	Y	Y	N	Y	N	N
Owning a subscription	Y	Y	Y	Y	Y	Y
Income level	Y	Y	Y	Y	Y	Y
<i>Objective</i>						
Trip character	Y	Y	Y	N	N	Y
Trip regularity	Y	Y	Y	Y	Y	N
Trip purpose	Y	N	N	N	N	N
Number of modes	Y	Y	Y	Y	Y	Y
Trip duration	Y	Y	Y	Y	Y	Y
Cost	Y	Y	Y	Y	Y	Y
Walking time	Y	Y	Y	Y	Y	Y
Waiting time	Y	Y	Y	Y	Y	Y
<i>Subjective</i>						
Satisfaction	Y	Y	Y	Y	Y	N
Convenience	Y	N	N	N	N	N
Feeling of control	Y	Y	Y	N	N	N
Perception as first choice	Y	Y	Y	Y	Y	N

BS1 represents *Backward Selection step 1*.

BS2 represents *Backward Selection step 2*.

TT1 represents *Threshold Test step 1*.

TT2 represents *Threshold Test step 2*.

HD represents *Hypotheses-Driven*.

Y indicates the factor is included in the set.

N indicates the factor is excluded from the set.

Backward selection with threshold The second strategy consisted of something similar to the first one. However, the removal criterion was more severe. Instead of removing factors whose coefficients showed no statistical significance for every single alternative, a threshold of three alternatives was set, meaning that a factor could only be considered relevant if its coefficients were, at a 5% significance level, significant for at least three modes. Take the example of *sex* after the first estimation: its coefficient showed statistical significance only for bus and metro; following this new criterion, it should be removed from the utility specification. In addition to *sex*, *age*, *trip character*, and *feeling of control* also fit this criterion. Such restriction resulted in the *Threshold Test step 1* set of factors (TT1), outlined in Table 3.9. A new and final iteration removed *household composition*, resulting in the *Threshold Test step 2* set of factors (TT2), as illustrated by Table 3.9.

Removing more than one factor at each estimation Note that, for its practicability and computation cost, the described strategies remove all the non-significant factors at each step, not only the least significant one, which is the most typical heuristic in a stepwise regression technique. In some cases, this more aggressive approach might overlook the interaction effect of some factors or result in the exclusion of factors with contextual importance, even if minimal.

Nonetheless, the methodology applied is highly reliable since it already accounts for three different selection strategies—even if they do not represent all possible combinations of factors. Likewise, mode choice is consistently independent of a few factors (*sex*, *convenience*, and *household composition*), meaning its exclusion is viable. Furthermore, even if removed, the inference on these and other factors is not that they are purely irrelevant but likely irrelevant (or only residually relevant) for most alternatives—at least for the estimated specifications.

Consequently, no contextual importance is compromised because no conclusions can derive from statistically insignificant factors, and no group of factors is ever wholly excluded. The only exception is the HD set, as seen in Table 3.9). Nevertheless, the study similarly did not derive any conclusions from the models built on it, as explained further.

Hypotheses-driven selection The third and final strategy builds on a series of hypotheses based on the population's context but also considers the statistical significance and computational effort observed after the initial estimation. The result is the most undersized set, HD (see Table 3.9), containing only 9 factors. Since this set did not yield satisfactory results for MLE's optimization algorithm while running it for the subsequent LCM, it was excluded from comparison to determine the best set of factors using the MNL approach, regardless of its validation results' quality.

3.3.2 Validation results

Tables 3.10, 3.11, and 3.12 show all the validation results for the assessment of the best set of factors when implementing MNL.

Table 3.10: In-sample validation results from MNL using all sets of factors, except HD.

Set	K	ρ^2	$\bar{\rho}^2$	$LL_e(\hat{\beta}^e)$	AIC	BIC	ACC	%CR	%CW	%U
Full	190	0.6255	0.2776	-204.5436	789.0872	1650.231	0.8632	67.54	3.93	28.53
BS 1	126	0.5855	0.3548	-226.3833	704.7667	1275.841	0.8559	63.46	4.37	32.17
BS 2	110	0.5840	0.3826	-227.1978	674.3956	1172.952	0.8544	63.90	4.51	31.59
TT 1	94	0.5645	0.3924	-237.8535	663.7069	1089.746	0.8457	63.61	4.08	32.31
TT 2	86	0.5630	0.4056	-238.6765	649.3531	1039.134	0.8501	62.88	3.93	33.19

For all sets of factors:

$$LL_e(0) = -546.1821$$

In-sample validation

All models see a significantly higher $\bar{\rho}^2$ in the ISV step, even if the reported ρ^2 is consistently lower compared to the baseline Full, as Table 3.10 shows. This trade-off means that, although model refinement reduces complexity and slightly sacrifices raw fit, it improves the model's generalization agility, likely preventing overfitting. For its robustness, and since it reflects a more realistic balance between model performance and complexity, the assessment of $\bar{\rho}^2$ was almost always deemed more important. In greater detail, the observed $\bar{\rho}^2$ positive differences between the sets corresponding to each step of the significance-driven strategies confirm the suspects regarding the factors removed and strongly point out that their removal was the adequate decision to achieve a more reasonable model. The $\bar{\rho}^2$'s improvement from 0.2776 (Full) to 0.4056 (TT2) is noteworthy. The lowest values for the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) also indicate better parsimony, so, in principle, the simplest model (TT2-derived model) should be favored over more complex ones.

The likelihood ratio test (LRT) was performed for all models using the baseline model (Full-derived) as the *unrestricted* model. Across all comparisons, the p-values are extremely high ($\sim 99.9\%$), indicating with an exceptionally high degree of confidence that all the simpler models do not significantly worsen the fit verified on the Full-derived model. The same test was repeated to ensure a complete comparison, as follows. First, it was applied to the models estimated on BS1 (as *unrestricted*) and BS2 (as *restricted*), indicating no significant differences. Secondly, the comparison between the models built on TT1 (as *unrestricted*) and TT2 (as *restricted*) had the same results. Thirdly and finally, since TT2 is *nested* within BS2, the Vuong test is not required for comparison, and the LRT was, again, performed—revealing the same outcome.

In terms of apparent predictive accuracy, there is a residual loss as the number of parameters decreases, resulting in a 1.31% ACC difference between the baseline model (Full-derived) and the TT2-derived model. There is also a slight decline in the latter's confidence in clearly correct predictions. This point illustrates a substitution pattern between %CR and %U, i.e., the model produces more predictions below the 0.75 probability threshold. Still concerning the TT2-derived model, although validation unveils such a slight confidence decrease for not being associated with the highest %CR, it shows that such a model possesses the lowest %CW (tied with the Full-derived model), which is of higher relevance.

Table 3.11: Out-of-sample validation results from MNL using all sets of factors, except HD.

Set	K	ρ^2	$\bar{\rho}^2$	$LL_t(\hat{\beta}^t)$	$LL_v(\hat{\beta}^t)$	ACC	PREC	F1
Training Data								
Full	190	0.6746	0.2494	-145.4096	-	-	-	-
BS 1	126	0.6273	0.3453	-166.5476	-	-	-	-
BS 2	110	0.6186	0.3724	-170.4262	-	-	-	-
TT 1	94	0.5880	0.3776	-184.1069	-	-	-	-
TT 2	86	0.5870	0.3946	-184.5306	-	-	-	-
Validation Data								
Full	190	-2.6772	-4.5901	-	-365.2298	0.8043	0.7953	0.7989
BS 1	126	-0.5855	-1.8541	-	-157.4831	0.7971	0.7889	0.7916
BS 2	110	-0.1471	-1.2546	-	-113.9349	0.7971	0.7897	0.7910
TT 1	94	0.1460	-0.8004	-	-84.8202	0.7899	0.7855	0.7839
TT 2	86	0.1365	-0.7293	-	-85.7644	0.7971	0.7938	0.7911

For all sets of factors:

$$LL_t(0) = -446.8581$$

$$LL_v(0) = -99.3240$$

Table 3.12: Validation sample's market shares (%) from MNL using all sets of factors, except HD.

Set	Bus	Metro	Own car	Walking	Ride-hailing	Ride	Bicycle/Scooter	Taxi	MAE
Market Shares - Observed									
-	34.06	33.33	14.49	12.32	2.17	2.17	1.45	0.00	-
Market Shares - Predicted									
Full	31.88	35.51	14.49	13.04	2.90	2.17	0.00	0.00	0.91
BS 1	31.16	36.23	14.49	13.04	2.90	2.17	0.00	0.00	1.09
BS 2	29.71	37.68	15.22	13.04	2.17	2.17	0.00	0.00	1.45
TT 1	28.99	39.13	13.77	13.04	2.17	2.90	0.00	0.00	1.81
TT 2	28.26	39.13	14.49	13.04	2.17	2.90	0.00	0.00	1.81

Ride represents *riding with someone*.

Out-of-sample validation

As Tables 3.11 and 3.12 demonstrate, the same conclusions as in ISV apply regarding good-ness-of-fit for training and validation data, with the TT2-derived model achieving the best results for $\bar{\rho}^2$. A critical remark is that for the models estimated using the second strategy of factor selection, the LL of the model estimated on the training data applied to the validation data is higher than the NLL of the validation data. This result indicates that these models perform better than the null model at explaining the observations of the validation sample and can capture some of its underlying patterns, suggesting higher robustness and reproducibility.

In this validation step, the LRT was only performed for the TT1-derived model, as the *unrestricted* model, compared to the TT2-derived model, as the *restricted* model. Section 3.5 briefly explains why such was done when evaluating a *restricted model with higher log-likelihood*. The results were the same as in ISV: the models are equally good regarding LL.

The accuracy of the validation data is tied to the second highest, performing slightly worse than the baseline model in $\sim 0.7\%$. The same applies to all the remaining classification performance metrics, meaning that the TT2-derived model has a comparable classification performance to the baseline. Finally, the predicted market shares show slightly worse alignment than the said model, corresponding to the tied worst MAE (see Table 3.12). This difference is also minimal ($\sim 0.9\%$), yet it must be seen cautiously, as it is underestimated since it represents a worsening of prediction capability regarding public transportation alternatives.

Overall analysis

The validation results of the four models estimated according to the two first factor selection strategies described in this section (BS1, BS2, TT1, and TT2) show a clear pattern of improvement, particularly in their final iterations. Similar enhancements appeared with the third strategy (HD). However, this study could not report its results, with the in-depth explanation for this impossibility presented in Section 3.5 when discussing the *computational issues* inherent to such a set of factors. Ultimately, the (system of) model(s) estimated on TT2 achieved the best performance, demonstrating robust performance across both validation steps. In ISV, this model achieves the highest adjusted $\bar{\rho}^2$ with a significant improvement from the Full-derived model (0.2776), alongside the lowest AIC and BIC values, indicating superior parsimony and a better balance of fit and complexity. It excelled in log-likelihood-related metrics in OSV, proving its overall robustness and suggesting it effectively captures the validation sample's underlying patterns. The negligible decrease in market share alignment, while employing 10 factors instead of 17, further ensures the hypothesis.

3.4 Nested and cross-nested logit models

The Nested Logit Model (NL), a widely used member of the Generalized Extreme Value (GEV) family, is an extension of the traditional MNL that divides the choice set C_n into subsets called

nests. This hierarchical structure allows alternatives within the same nest m to have correlated error terms ε_{in} . The model accounts for similarities between alternatives within the same nest by relaxing the independence of irrelevant alternatives (IIA) assumption. Building on the definitions inherent to MNL, the probability of choosing an alternative i is given by Eq. 3.13, where the probability $P_{i|m}$ of choosing alternative i within nest m is computed by Eq. 3.14 and the probability $P_{m|C_n}$ of choosing nest m is computed by Eq. 3.15.

$$P_{i|C_n} = P_{i|m} \cdot P_{m|C_n} \quad (3.13)$$

$$P_{i|m} = \frac{e^{\mu_m V_{in}}}{\sum_{j \in m} e^{\mu_m V_{jn}}} \quad (3.14)$$

$$P_{m|C_n} = \frac{e^{\lambda_m I_m}}{\sum_{k \in C_n} e^{\lambda_k I_k}}, \quad \text{where } I_m = \ln \left(\sum_{j \in m} e^{\mu_m V_{jn}} \right) \quad (3.15)$$

In Eq. 3.14, μ_m is the scale parameter for alternatives within a nest, determining the degree of correlation among alternatives in the same nest. Smaller values of μ_m indicate a stronger correlation among alternatives within the nest, while $\mu_m = 1$ represents no correlation (independence) within the nest, as in MNL. Similarly, λ_m is the scale parameter for nests, controlling the level of independence between nests. The NL model collapses to MNL when $\mu_m = \lambda_m = 1$ for all nests, indicating no correlation among alternatives within or across nests. Finally, I_m , the log-sum term, captures the overall attractiveness of the nest m by aggregating the utilities of all alternatives in that nest.

The Cross-Nested Logit Model (CNL) extends NL by allowing alternatives to belong to multiple nests, with fractional allocations to each nest. This novel feature increases flexibility in capturing substitution patterns and cross-elasticities between alternatives. In this configuration, $\alpha_{i,m}$ is the fractional allocation of alternative i to nest m and λ_m and I_m are defined similarly as in NL. Eq. 3.16 expresses the choice probability in CNL as follows:

$$P(i|C_n) = \frac{e^{V_{in}} \sum_m (\alpha_{i,m} e^{\lambda_m I_m})}{\sum_{j \in C_n} e^{V_{jn}} \sum_n (\alpha_{j,n} e^{\lambda_n I_n})}, \quad \text{where } \sum_m \alpha_{i,m} = 1 \text{ for all alternatives } i. \quad (3.16)$$

The present study implemented these two models, experimenting with various nests, overlapping nests, and different predefined allocation weights in the specific case of CNL. Several utility specifications were also proposed. Despite the variety of trials for both model types, the yielded results consistently failed to meet expectations, as at least one of the following issues always arose: no convergence of the optimization algorithm for the parameter estimation; parameters showing low to no statistical significance, leaving no room for practical interpretation; costly computation. Nevertheless, although various estimated parameters did not correspond to the maximum likelihood estimate, the validation results mainly showed somewhat comparable results to MNL. Both

facts support the evidence from the literature described in Section 2.2.4. Likewise, as discussed in Section 2.2.5, a valid hypothesis for one of the many root causes of the problem is that the many alternatives might have had a crucial role in such estimation results, given the high complexity they eventually imply. All in all, the results underscore the practical requirements of such methodologies. A more detailed description of the various modeling attempts is as follows.

Nested logit The first group of nests proposed for NL, given the most intuitive similarities between studied alternatives, included *public transportation* (bus, metro), *ride services* (taxi, ride-hailing), *personal transportation* (own car, riding with someone), and *non-motorized* (bicycle/scooter, walking). Other configuration only included *friendly modes* (bus, metro, bicycle/scooter, walking) and *car modes* (taxi, ride-hailing, own car, riding with someone). A third experiment contemplated *personal transportation* (own car, riding with someone) and *non-personal transportation* (bus, metro, taxi, ride-hailing, bicycle/scooter, walking). The estimation for any of the presented scenarios was computationally challenging. Likewise, it did not result in the convergence of MLE's optimization algorithm, regardless of the set of factors employed and also independently of the validation step.

The study further implemented several approaches to resolve this issue. In terms of included factors, it employed a *forward selection* technique, starting by only specifying the ASCs for all alternatives and adding each factor at a time. First was cost, followed by trip duration, and, thirdly, a factor from a different group. The previously described limitation arose when including this third factor, regardless of its nature (e.g., satisfaction, age, or income level). A second approach consisted of normalizing the whole working data, yet the problem persisted despite a higher-speed estimation. The optimization algorithm's maximum number of iterations was augmented in a third attempt. This approach yielded no results, as the number of iterations never reached the maximum, unveiling that it was not a cause of termination of the algorithm. Another optimization algorithm enabled by Biogeme was selected but also provided no improvements.

Cross-nested logit In the context of CNL, the study began by overlapping nests in the following manner: *public transportation* (bus, metro), *personal transportation* (taxi, ride-hailing, on car, riding with someone, bicycle/scooter, walking), *motorized* (bus, metro, taxi, ride-hailing, own car, ride) and *non-motorized* (bicycle/scooter, walking). In an initial estimation with the Full set, this framework yielded unsatisfactory results, as in NL. However, the most negative aspect emphasized was the highly extended period the estimation process endured, leading to abandoning it while it was being performed. This study attempted to estimate the factors in TT1 under the exact nest framework described above. In this novel attempt, the critical issue was the lack of statistical significance (all p-values equal to zero) in the estimation to fit the training data. Given the improvements in the estimation in ISV, TT1 was once more used to estimate the same overlapping nests as previously, with a fifth nest: *ride services* (taxi, ride-hailing). Given the arguably higher complexity introduced by the new nest, the new estimation of the training data did not produce any results, significant or not, due to the posed time restrictions. Other approaches included a

combination of: the addition of a *friendly* (bus, metro, bicycle/scooter, walk) and a *non-friendly* (taxi, ride-hailing, own car, riding with someone) nests; the determinist assignment of the allocation parameters $\alpha_{i,m}$ for some alternatives, in a an attempt to optimize the estimation procedure; the implementation of the remaining sets of factors proposed by the study. Persistent limitations in this myriad of options included the exact ones found in prior attempts, with a slightly more stable estimation of ISV. It is likewise hypothesized that such an algorithm could be sensitive to data quality and that *hiding* valuable data points for OSV's validation sample might have a crucial role in manifesting such difficulties.

3.5 Latent class model

A Latent Class Model (LCM) assumes heterogeneity of preferences—contrasting to the homogeneity inherent to MNL—by diving the population into the so-called *latent classes*, where individuals belonging to each class possess distinct preference structures. This assumption of LCM translates to the following: individual behavior can depend on observable factors and unobservable heterogeneity. Such property constitutes a substantial advantage over MNL since this model assumes a single preference structure applied uniformly to the entire population. Substitution patterns in LCM are also unrestricted globally, although still applicable within each class, where the assumed model is an MNL (Train, 2009d). This model is, hence, a special case of the Mixed Logit Model (MMNL) (McFadden and Train, 2000), where a discrete distribution substitutes the continuous distribution assumed for each coefficient of the MMNL (Greene and Hensher, 2003). Nonetheless, the class membership is not directly observable. Each individual n has a probability H_{nq} of belonging to each class q rather than being deterministically assigned to a specific class among the Q classes. This probability can be computed similarly to the choice probability of MNL, as in Eq. 3.17. In such formulation, the vector of specified attributes x'_n of class q is multiplied by its β_q coefficients, thus representing the systematic utility V_{nq} of any individual belonging to class q .

$$H_{nq} = \frac{e^{x'_n \beta_q}}{\sum_{r=1}^Q e^{x'_n \beta_r}} = \frac{e^{V_{nq}}}{\sum_{r=1}^Q e^{V_{nr}}} \quad (3.17)$$

Within each class, the choice probabilities are consistent with the logit structure since MNL is the reference model of this study. Therefore, they are formulated according to Eq. 3.5. Considering the latent class membership, the overall probability of choice regarding alternative i concerning individual n is given by the mixture of class-specific probabilities:

$$P_{in} = \sum_{q=1}^Q H_{nq} \cdot P_{in|q} \quad (3.18)$$

This study proposes dividing the population into two segments, thus presenting an LCM with two classes ($Q = 2$). Preliminary computational experiments have shown that the software used

Table 3.13: In-sample validation results from LCM using sets of factors BS2 and TS2.

Set	K	ρ^2	$\bar{\rho}^2$	$LL_e(\hat{\beta}^e)$	AIC	BIC	ACC	%CR	%CW	%U
BS 2	221	0.6582	0.2536	-186.6674	815.3348	1816.981	0.8806	60.84	2.18	36.97
TT 2	173	0.6999	0.3831	-163.9295	673.8591	1457.953	0.8588	64.48	0.00	35.52

For both sets of factors:

$$LL_e(0) = -546.1821$$

for estimation could not handle more classes for these specific models (with a high number of factors), unveiling the computational limitations inherent to this different approach. Nevertheless, it has been shown that a high number of classes does not necessarily provide better results (Greene and Hensher, 2003), and two classes can be considered an adequate starting point to analyze the underlying heterogeneity in the pool of respondents. Therefore, in this LCM application, Eq. 3.18 can be specifically reduced to:

$$P_{in} = H_{n1} \cdot P_{in|1} + (1 - H_{n1}) \cdot P_{in|2} \quad (3.19)$$

Although different utility specifications are possible for each latent class, the proposed LCM presents the exact specification for its two classes under the same behavioral assumption of MNL, as in Eq. 3.6. Additionally, both classes reflect the addition or removal of any specified factors: e.g., if there is a need to remove cost from the utility specification, such is reflected in both classes. In practical terms, what this means is that, for the same set of factors, MNL and LCM, together, present three distinct coefficient results for the same utility function, plus an additional coefficient corresponding to the latent class probability of any individual belonging to class 1 (H_{n1}).

3.5.1 Validation results

There is model improvement in terms of validation results and estimation efficiency between the first (BS1 and TT1) and last (BS2 and TT2) iterations of the significance-driven factor selection strategies described in Section 3.3. Estimating the parameters for LCM is also a more computationally expensive process. Based on this, the sets corresponding to the first iteration of each strategy were excluded from the LCM parameter estimation to achieve greater confidence in improved results and efficiency. Tables 3.13, 3.14, and 3.15 display the results of such estimations using the said sets of factors (BS2 and TT2).

In-sample validation

In terms of ISV, the TT2-derived model performed better on almost all metrics (see Table 3.13), except accuracy, which is marginally lower but still considerably high (85.88%). Likewise, it achieves a remarkable %CW of 0.00% for the predefined threshold, meaning it makes no confident predictions above 75% alternative probability that are incorrect. This null percentage reflects the model's higher reliability in its predictions, mitigating its slightly lower apparent predictive power.

Table 3.14: Out-of-sample validation results from LCM using sets of factors BS2 and TT2.

Set	K	ρ^2	$\bar{\rho}^2$	$LL_t(\hat{\beta}^t)$	$LL_v(\hat{\beta}^t)$	ACC	PREC	F1
Training Data								
BS2	221	0.7580	0.2634	-108.1424	-	-	-	-
TT2	173	0.6719	0.2847	-146.6231	-	-	-	-
Validation Data								
BS2	221	-26.6333	-28.8583	-	-2744.6500	0.7536	0.7607	0.7465
TT2	173	-8.4924	-10.2342	-	-942.8250	0.7319	0.7313	0.7289

For both sets of factors:

$$LL_t(0) = -446.8581$$

$$LL_v(0) = -99.3240$$

The $\bar{\rho}^2$ is substantially higher (0.3831 vs. 0.2536), and the AIC-BIC pair presents the lowest values, suggesting higher model parsimony. It achieved a better fit even with accounting for complexity, as denoted by the ρ^2 results (0.6999 vs. 0.6582). Although the BS2-derived model has a marginally higher accuracy (2.18%), the difference becomes negligible when accounting for the remaining performance improvements of the competing model.

Restricted model with higher log likelihood Since TT2 is a special case of BS2, the LRT statistic could be computed using their respective models as *restricted* and *unrestricted* models. However, the *restricted* model (TT2-derived) has a LL ($LL_{restricted}$) closer to zero, making $LL_{unrestricted} - LL_{restricted}$ a negative difference, which is an unexpected computation anomaly in the context of the chi-squared distribution. Such an anomaly, per se, reaffirms that the parameter reduction does not significantly worsen the model's fit; instead, it has the opposite effect. For the same reason, the same test was not performed on OSV for the LLs obtained with the validation data.

Table 3.15: Validation sample's market shares (%) from LCM using sets of factors BS2 and TT2.

Set	Bus	Metro	Own car	Walking	Ride-hailing	Ride	Bicycle/Scooter	Taxi	MAE
Market Shares - Observed									
-	34.06	33.33	14.49	12.32	2.17	2.17	1.45	0.00	-
Market Shares - Predicted									
BS2	27.54	42.03	10.14	13.77	3.62	2.90	0.00	0.00	3.08
TT2	31.88	36.23	12.32	13.04	2.17	4.35	0.00	0.00	1.45

Ride represents *riding with someone*.

Out-of-sample validation

The system of models built on TT2 excelled with superior $\bar{\rho}^2$ for both training and validation data, even though both models show negative values for the latter (see Table 3.14). Both models also show similar drops in terms of accuracy ($\sim 12.7\%$) when moving from ISV to OSV, indicating a potential degree of overfitting.

When performing the LRT to compare the LLs in the context of the training data (TT2-derived model as *restricted*, BS2-derived model as *unrestricted*), the LRT value corresponded to a p-value of $\sim 0.5\%$. The null hypothesis was rejected, indicating that the *unrestricted* model provides a significantly better fit to the data than the *restricted* model. Given the consistency of previous results and estimations, such a result seemed counterintuitive and resulted in further inspection, culminating in resorting to AIC and BIC on OSV. These criteria values are still smaller for TT2 (~ 658 vs. ~ 639 and ~ 1610 vs. ~ 1385 , respectively), as in ISV, which also aligns with the $\bar{\rho}^2$ value. Even if it does not provide a similar raw fit when estimating the model in the training sample, the simplest model (built on TT2) provides the best balance of fit and complexity.

Moreover, in alignment with the validation ideology, both models' classification performance and aggregate outcomes should be critically evaluated, avoiding excessive reliance on fit metrics alone. The BS2-derived model again shows better classification performance with narrowly superior results across all the related metrics. Nonetheless, in congruence with the flawed goodness-of-fit results for the validation data, these metrics degrade substantially for both models, as previously exemplified by the ACC drop.

The opposing set compensates with better market share alignment through lower MAE (1.45% vs. 3.08%), but also by only failing to best predict individuals who ride with someone (see Table 3.15). In this context, note that the BS2-derived model highly overestimates the share for metro (27.54% predicted vs. 34.06% observed) and also underestimates the share for bus (42.03% predicted vs. 33.33% observed). These discrepancies underscore the robustness *system of models* estimated on TT2 in capturing market share distributions more accurately, particularly in essential cases where the other exhibits instability.

Computational issues—optimization algorithm non-convergence As stated in Section 3.3, HD was also used for the LCM implementation, however, resulting in no convergence of the optimization algorithm while estimating the model on the training data for OSV validation. This situation had been verified before while using the dataset generated with RFI, as mentioned in Section 3.2, and many times during the several trials using the NL and CNL extensions, as described in Section 3.4. This is problematic because it indicates that the estimated parameters do not correspond to the maximum likelihood estimate since the algorithm fails to find the optimal values of the parameters that maximize the likelihood function. In this sense, the estimates likely do not reflect the actual underlying patterns and inherent behavior expressed in the data, and the reliability of the conclusions derived from their interpretation is low. Such an issue is typically accompanied by poor statistical inference, as the values of the parameters' standard deviations and t-statistics/p-values are extremely large or small. For example, one of the parameters possesses a

standard deviation of 139 units and is associated with a parameter estimated at an absolute value of 0.188 units, which is disproportionate and naturally associated with a t-statistic close to zero, making parameter interpretation impossible. Along the same lines as what happened with the nested extensions (NL and CNL), the validation did not degrade per se, yet, for the outlined reasons, the comparison cannot be made. This issue emphasizes the need to address convergence problems and carefully inspect beyond validation measures.

Overall analysis

The TT2-derived model demonstrates superior performance across both ISV and OSV. In ISV, it achieves a higher adjusted $\bar{\rho}^2$ (0.3831 vs. 0.2536) and lower AIC and BIC values, indicating a better balance between model fit and complexity. Additionally, it performs no confident prediction errors (%CW=0.00%), highlighting its reliability. In OSV, the TT2-derived model maintains its robustness with a better $\bar{\rho}^2$ for both training and validation datasets and a lower MAE of 1.45% in market share alignment compared to 3.08% for the BS2-derived model. These results and its computational stability and avoidance of convergence issues establish the TT2-derived model as this study's most reliable LCM choice.

Chapter 4

Results and discussion

When evaluating a DCM, the last step is to interpret the estimated coefficients and fully deploy the resulting insights and implications. After a comparative assessment of model performance, conducted in Section 4.1 in uniformity with what is proposed throughout the study, these findings must be contextualized within the specific environment where the study was conducted. Afterward, Section 4.2 provides a detailed interpretation of the estimation results for MNL and LCM.

4.1 Validation results: multinomial logit vs. latent class

Special caution must be exercised when comparing the validation results from models belonging to distinct variants. This is the current case for MNL and LCM, estimated on TT2. Tables 4.1, 4.2, and 4.3 display the validation results, putting to evidence that, while MNL presents superior performance results in key areas, it may have distinct restrictions and limitations not as perceivable at first sight, as previously mentioned. Such happens because these limitations are mostly related to the nature of the estimated coefficients and the way they reflect real-world patterns. Likewise, the fact that one can analyze systematic taste heterogeneity among classes using LCM draws considerably more attention to such a model. However, the model in question is also more complex. Thus, in relative contrast to what this study previously presented, the purpose of this validation process is, ultimately, not to exclude but to observe the advantages and limitations of each model using the reported metrics. Such a trade-off and methodological assessment must not be disregarded when interpreting both model's coefficients.

Table 4.1: In-sample validation results from LCM and MNL using set of factors TT2.

Model	K	ρ^2	$\bar{\rho}^2$	$LL_e(\hat{\beta}^e)$	AIC	BIC	ACC	%CR	%CW	%U
LCM	173	0.6999	0.3831	-163.9295	673.8591	1457.953	0.8588	64.48	0.00	35.52
MNL	86	0.5630	0.4056	-238.6765	649.3531	1039.134	0.8501	62.88	3.93	33.19

For both models:
 $LL_e(0) = -546.1821$

Table 4.2: Out-of-sample validation results from LCM and MNL using set of factors TT2.

Model	K	ρ^2	$\bar{\rho}^2$	$LL_t(\hat{\beta}^t)$	$LL_v(\hat{\beta}^t)$	ACC	PREC	F1
Training Data								
LCM	173	0.6719	0.2847	-146.6231	-	-	-	-
MNL	86	0.5870	0.3946	-184.5306	-	-	-	-
Validation Data								
LCM	173	-8.4924	-10.2342	-	-942.8250	0.7319	0.7313	0.7289
MNL	86	0.1365	-0.7293	-	-85.7644	0.7971	0.7938	0.7911

For both models:

$$LL_t(0) = -446.8581$$

$$LL_v(0) = -99.3240$$

In-sample validation

Table 4.1 demonstrated that, on the one hand, MNL presents the best balance of fit and complexity: naturally, since both models use the exact utility specification consistently, the MNL possesses half the parameters present in LCM, minus H_{n1} , i.e., the probability associated with belonging to LCM's Class 1 (LCM1); its $\bar{\rho}^2$ is slightly higher (0.4056 vs. 0.3831); AIC and BIC are lower (see Table 4.1).

Theoretically, an LCM can be reduced to MNL when the number of latent classes collapses to one, i.e., when class probabilities become deterministic. In this case, both models are estimated on the same set of attributes and share the exact utility specification. Hence, the consideration of MNL as a special case of LCM, as stated by Greene and Hensher (2003), can be made. LRT was, therefore, applied and resulted in a p-value exceptionally close to zero (0), showing strong evidence supporting the hypothesis that, on the current application, with the same data and utility specification, LCM provides a better fit to all observations. The result undoubtedly results from increased model flexibility, but it can also arguably prove the model's tendency to overfit.

On the other hand, Table 4.1 also shows that LCM slightly overcomes MNL in terms of apparent predictive accuracy (ACC = 0.8588 vs. 0.8501), with both models achieving excellent results. Additionally, LCM makes predictions that are significantly more confident. It is not only better in confidently predicting correctly (%CR = 64.48% vs. 62.88%), but it also does not make wrong predictions based on the alternative's probability above 0.75 (%CR = 0.00%). This noteworthy result must be considered when evaluating OSV's classification and market share results.

Out-of-sample validation

This study finds a significantly similar pattern to ISV regarding the model's goodness-of-fit to OSV training data, as illustrated by 4.2. There are, however, two caveats. First, the $\bar{\rho}^2$ difference becomes substantially more pronounced (0.2847 vs. 0.3946). Second, when evaluating the LL, LRT informs that the complex model does not significantly improve over the simpler model. Thus, as in ISV, MNL provides the best balance of fit and complexity for the training data while

Table 4.3: Validation sample's market shares (%) from LCM and MNL using set of factors TT2.

Set	Bus	Metro	Own car	Walking	Ride-hailing	Ride	Bicycle/Scooter	Taxi	MAE
Market Shares - Observed									
-	34.06	33.33	14.49	12.32	2.17	2.17	1.45	0.00	-
Market Shares - Predicted									
LCM	31.88	36.23	12.32	13.04	2.17	4.35	0.00	0.00	1.45
MNL	28.26	39.13	14.49	13.04	2.17	2.90	0.00	0.00	1.81

Ride represents riding with someone.

also providing equally good results in terms of fit. When analyzing the same metrics for the validation data, conversely to previous scenarios, the complex model achieves a lower LL than the simpler model. Only the MNL can better explain the validation sample than the null model. This reproducibility capacity is noteworthy, and LCM's tendency to overfit, as discussed in Section 3.5, renders the comparison considerably more compelling and conclusive.

The pronounced reduction in ACC from ISV to OSV ($\sim 12.7\%$ vs. $\sim 5.3\%$) reflects LCM's overfit propensity again. This drop is accompanied by a consistent underperformance in all classification metrics by $\sim 6.3\%$, on average, as shown by Table 4.2.

Surprisingly, notwithstanding being a worse classifier at the individual level, when aggregating predictions, LCM performs comparably to MNL (MAE = 1.45% vs. 1.81%), as shown by Table 4.3. An argument could be made that LCM is slightly better, actually: while both models predict the same proportions for taxi, ride-hailing, bicycle/scooter, and walking, MNL is more inclined to make optimal predictions for car alternatives (own car and ride), and LCM is most aligned with public transportation (bus and metro).

Overall analysis

LCM is better suited for analyses requiring nuanced subgroup-specific insights, whereas MNL is more robust for applications where simplicity and generalizability are critical. In ISV, MNL demonstrates better parsimony despite having half the parameters of LCM. However, LCM slightly outperforms MNL regarding predictive accuracy, making no clearly wrong predictions above the confidence threshold. MNL further establishes its robustness in OSV by achieving better reproducibility on the validation sample. Although it surprisingly achieves the best aggregate outcomes, LCM seems limited by a higher proneness to overfit.

4.2 Estimation results and analysis

The parameter estimation results from MNL are reported in detail in Table 4.4, as an illustration, and suggestion to explore the remaining results concerning LCM Tables B.1, and B.2 of Appendix B. This section discusses the results yielded by both models simultaneously to avoid excessive repetition and facilitate a more comprehensive interpretation of each factor and/or parameter. All

the results and derived inferences must be read as part of the specific context in which the study was conducted. In regards to LCM, the likelihood of any individual possessing the preference patterns reflected by the coefficients estimated for LCM's Class 2 (LCM2) is higher ($H_{n2} = 0.603 > H_{n1} = 0.397$). This class imbalance is considered when performing a global evaluation of each factor.

4.2.1 Alternative-specific constants

For the ASCs, the inferences derived from the presented parameter values assume *ceteris paribus* ("all other things being equal").

Summary of main insights: In the MNL, the systematic preference for any given mode over bus (reference alternative) is significant only for own car. Globally, when considering the LCM, the hierarchy of systematic preference for the population in this study can be inferred, albeit with some nuances, as follows in descending order: own car, metro, riding with someone, walking, bus, and ride-hailing. However, the results for taxi and bicycle/scooter are mainly inconclusive.

Own car

The mentioned effect in MNL is associated with the highest coefficient for any ASCs. Moving to LCM, such an effect is only significant for LCM2, corresponding again to the strongest systematic preference of any mode over bus. In LCM1—although the class with lower likelihood—in addition to displaying no statistical significance, the ASC has a minimal, almost irrelevant, magnitude. Consequently, this study concludes that there is a firm preference for own car over bus for most of the studied population. This finding applies not only to bus, but to all the alternatives compared to own car by inference, for the following reasons.

Metro

On LCM, this mode is consistently preferred over bus in both classes, displaying an even more substantial relative magnitude in LCM2. In this class, such preference is also the second most notable one. With MNL, nonetheless, the differences, although pointing to the same direction ($ASC_{metro} > 0$), are not substantially evident and not even statistically significant. Overall, the results suggest metro is preferred to bus.

Concerning own car, the only comparison where the ASC of both alternatives shows statistical significance is inside LCM2. Here, ASC_{OwnCar} is almost twice the value estimated for metro, even though the latter is the second highest. An even more substantial effect is verified in MNL, though reversed in LCM1. The results may be partially inconclusive due to a lack of statistical significance. However, at least a considerable portion of the population is most likely to prefer own car over metro. It is highly reasonable to consider own car more preferable to metro.

Table 4.4: Parameter estimation results from MNL using set of factors "Threshold test step 2".

Factor	Bus (N = 240)	Metro (N = 220)	Walk (N = 98)	Own Car (N = 93)	Riding with someone (N = 16)	Ride-hailing (N = 15)	Bicycle/Scooter (N = 3)	Taxi (N = 2)
ASC	0.00 ^a —	0.74 (0.46)	3.82 (0.89)	12.70*** (3.96)	2.82 (0.67)	-4.95 (-1.09)	-0.49 (-0.07)	3.25 (0.36)
<i>Socio-demographic</i>								
Owens a subscription: yes	3.24*** (5.66)	2.11*** (3.47)	-10.00*** (-8.90)	0.99 (1.44)	1.71* (1.65)	1.86** (2.24)	-1.49 (-0.79)	1.59 (1.50)
Owens a subscription: no ^b	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Income level: high	-3.17*** (-4.98)	-3.14*** (-4.64)	2.99*** (3.07)	-1.70** (-2.17)	1.57 (1.02)	-1.67 (-1.46)	3.39** (2.28)	1.73 (1.31)
Income level: low	-1.41*** (-2.77)	-1.46*** (-2.74)	2.17*** (3.11)	-0.44 (-0.62)	-0.79 (-0.97)	-0.03 (-0.02)	0.21 (0.17)	1.75 (0.69)
Income level: none ^b	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>Objective</i>								
Trip regularity	0.65*** (3.06)	1.06*** (4.65)	1.23*** (4.45)	1.04*** (3.67)	1.19*** (3.14)	0.90*** (3.12)	-3.13*** (-3.88)	-2.95*** (-4.63)
Number of modes	14.40*** (12.30)	15.40*** (13.20)	—	6.88*** (5.52)	14.70*** (11.80)	16.10*** (12.20)	14.10*** (11.90)	5.71 (1.31)
Trip duration	0.33*** (5.00)	0.24*** (3.39)	0.87*** (9.85)	0.20** (2.05)	0.24 (1.63)	0.25** (2.26)	-0.19 (-0.73)	-1.94*** (-5.97)
Cost	5.54*** (7.51)	5.33*** (7.34)	—	5.66*** (7.44)	5.61*** (7.52)	5.87*** (7.60)	1.08 (0.78)	7.84*** (9.05)
Walking time	0.32*** (2.77)	0.37*** (3.23)	—	0.11 (0.83)	-0.36 (-1.54)	—	-5.65*** (-7.44)	—
Waiting time	1.11*** (4.31)	0.97*** (3.80)	—	—	—	0.95*** (2.98)	—	0.91*** (3.85)
<i>Subjective</i>								
Satisfaction	-1.78*** (-3.45)	-1.81*** (-3.47)	-2.27*** (-2.78)	-1.62*** (-3.06)	-0.77 (-1.16)	-0.38 (-0.43)	4.03** (2.25)	4.59** (2.10)
Perception as first choice	0.01 (0.03)	0.00 (0.01)	2.94*** (7.48)	-0.01 (-0.03)	-0.77* (-1.72)	-0.84** (-2.20)	1.52 (1.11)	-2.85*** (-4.12)

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$ t -statistics in parentheses.

"Ride" stands for riding with someone.

^a Bus is the reference alternative, therefore its constant is normalized to zero.^b Reference category for the respective factor, therefore all coefficients are normalized to zero.^c Not applicable for the respective alternative.

Walk

In regards to walking, the preference pattern shares some similarities with metro. This similarity also derives from the results of LCM, where, nonetheless, the magnitude of the coefficients is substantially smaller, being close to zero for LCM1. In MNL, this preference is the second strongest positive effect, though it is not statistically significant. Regardless, there is enough information to state that most individuals are likely to prefer walking over traveling by bus. Even though $ASC_{Metro} < ASC_{Walk}$ in MNL, the comparison is somewhat inconclusive due to the lack of statistical significance for both coefficients. Thus, with the reasons stated in the comparison of walking to bus, it must be inferred that metro is systematically preferred over walking.

Ride with someone

The coefficients for riding with someone are similar to those obtained for walking. Hence, similar conclusions can be drawn. When comparing these two modes, though, a case could be made for a slight preference for riding with someone over walking, as seen in LCM2, where the difference in favor of riding with someone is more noticeable than the one verified in LCM1 in favor of walking. MNL presented, again, statistically inconclusive results, though walking has the edge in this case.

Ride-hailing

Ride-hailing is the least preferred alternative in any of the cases. In absolute value, this lack of preference is the second most substantial effect registered with MNL—though not significant—and LCM1. This evidence suggests that individuals consistently avoid using this alternative.

Bicycle/scooter

In the case of bicycle/scooter, while its ASC has a positive sign for LCM1, the effect inverts in LCM2, with the same happening in MNL. Nonetheless, the latter is not statistically significant, and all the registered magnitudes of this coefficient are close to zero, displaying even lower relative relevance in the context of LCM, where magnitudes for other alternatives are considerably greater. As with taxi, the results do not represent a substantial sample of the population—the support is 3 and 2 individuals for each, as indicated (see, e.g., Table 4.4). According to the available information, though, bicycle/scooter appears to share a similar preference level to bus, with a slight advantage for bus.

Taxi

Although MNL and LCM2 display a positive ASC, such values have no statistical relevance. In LCM2, this coefficient is the closest to zero verified in the presented estimations, making it potentially as preferable as a bus for a portion of the population. These results are primarily inconclusive and require further investigation, instead with a more significant sample of individuals who chose this mode. Such is, again, applicable to bicycle/scooter.

4.2.2 Socio-demographic factors

Summary of main insights: Though only two socio-demographic factors were explicitly studied, interesting findings follow: (1) owning a subscription strongly promotes public transportation while mostly discouraging walking, and (2) higher incomes generally decrease the likelihood of choosing public transportation while positively impacting walking and private modes. Age and sex are highly likely not impactful factors in this context, and household composition, although of likely limited relevance, merits further investigation. A detailed analysis of these insights follows.

Owning a subscription

According to the coefficients estimated via MNL, individuals who own any subscription to access public transportation are naturally more likely to use this mode type. This result is supported by a notably high statistical significance, with bus having a slightly higher coefficient than metro. In fact, according to MNL, individuals who choose bus yield the highest utility from owning a subscription. The scenario changes slightly in LCM, with metro presenting a negative coefficient for this factor in LCM1—although this is the class associated with the preference patterns where metro possesses the highest systematic preference. Inside this group, individuals owning a subscription are typically more likely to opt for bus to the detriment of metro when faced with both public transportation alternatives. LCM2 confirms the same result observed in MNL, with metro exhibiting a substantially stronger coefficient (nearly three times that of bus). Since LCM1 probabilistically represents a lesser portion of the overall population, it is still reasonable to extend, in general, the conclusions from MNL to LCM regarding public transportation as a whole.

For the three examined sets of coefficients (from MNL, LCM1, and LCM2), walking is the alternative with the most substantial negative coefficient for owning a subscription. All these coefficients are associated with an extremely high statistical significance; therefore, this study has enough evidence to state that people who own a subscription are less likely to walk among all the presented alternatives. Of those who own a subscription, only 16.09% did not pay for it. For the remaining individuals, the average subscription cost is ~30.19 euros, meaning that, for most subscription owners, owning a subscription represents a fixed cost already incurred in exchange for unlimited or discounted use of public transportation. Maximizing this value implies opting more frequently for this mode type, to the detriment of, e.g., walking. This scenario is potentially more interesting for the individual, mainly when public transportation offers lower travel time than walking to the same destination. Thus, the effect reflected on the coefficient results for walking might happen because this subscription encourages them to use public transportation instead, making the trip more efficient and cost-effective.

The conclusions verified for walking in the LCM can also be applied to own car and bicycle/scooter. However, the presented coefficient magnitudes are lower, particularly for bicycle/scooter. The coefficient estimated in MNL is associated with no statistical significance for either alternative.

Individuals who chose to ride with someone also see a positive utility impact from owning a subscription across all scenarios—with the coefficient for MNL being significant only at a 10% level and the magnitude of coefficients in LCM being lower than those of public transportation. 75.00% of these individuals own a subscription, and, of those, 66.67% paid for it. Additionally, the data on number of modes (used in a total journey) show that bus, ride with someone, and metro, excluding low support alternatives, have the highest averages for the given factor. Therefore, a plausible explanation for the results is that there might be a tendency to combine any of these three modes for higher convenience and flexibility. Also, individuals who rode with someone might occasionally use public transportation themselves.

Although MNL resulted in a significant negative coefficient for ride-hailing, LCM points to inconclusive results (coefficients with opposite signs), displaying the heterogeneity among those who chose this mode. For taxi, the results are, once more, also inconclusive.

Income level

In the MNL, having any form of income, whether high or low, has a negative impact on the utility yielded from choosing public transportation. The effect is notably more substantial for the highest income level. Moreover, the bus and metro coefficients are almost the same for this factor. A similar behavioral pattern is verified in LCM, though with slight caveats, and displaying insights is only possible due to the latent class approach. In LCM1, there is a positive utility associated with metro for low-income individuals, an effect reversed in LCM2 and exhibiting a considerably greater magnitude than that observed for bus—whose coefficient is non-significant. Still, regardless of income level, given its high systematic preference in this class, individuals are most likely to prefer metro over bus. For both classes, metro's coefficient for high-income level, although still negative, has a lower magnitude than bus'. Consequently, the present study finds that higher income levels are associated with a lower likelihood of choosing public transportation, which is even more evident in the case of bus. Although not as impactful, having some form of income also seems to diminish the likelihood of choosing these alternatives, particularly for bus.

Walking is consistently preferred for individuals with income, regardless of the approach used. Nonetheless, the different approaches show disparities in results for high and low-income levels: even if all coefficients are positive, MNL suggests high-income individuals are more likely to walk than low-income individuals. In contrast, LCM shows the opposite in both classes. According to the data, walking is the most chosen mode for leisure trips (39.74%), and this purpose is the second most popular among individuals choosing to walk (31.63%), only lower than study (39.80%) by a relatively small margin. A vast majority of leisure travelers (82.05%) are also known to report some form of income. This effect may indirectly reflect the propensity of individuals reporting income to walk for leisure.

Results for own car promote a more in-depth discussion and, once again, reveal the power of the latent class approach in evaluating population heterogeneity. According to MNL, high-income levels are significantly associated with a lower likelihood of driving. The model is also

inconclusive in regards to low-income level for this alternative. Moving to LCM, the total opposite is expected for individuals sharing the preferences of LCM1, which is also inconclusive for low-income in this class for this alternative. Nonetheless, in LCM2, while no conclusions can be drawn from the high-income category, low income is significantly associated with a lower utility from driving. MNL might show discrepancies with the literature, but LCM can better distinguish the differences in terms of income level: the present study concludes that a significant share of individuals possessing a high-income level are more likely to choose to drive, while another portion of individuals with a low income yields less personal satisfaction, i.e., utility from driving. Low-income individuals may prefer walking over own car, especially for shorter trips, to avoid traffic or the nuisance that parking often represents. Moreover, the data inform that own car is the most chosen mode for hospital visits (36.26%), but also that hospital visit is the most common purpose for driving (35.48%) after commuting to work (29.03%). For these trips, choosing to drive could have been seen by some individuals as a dis-utility a priori, even for those who report low income. However, they may choose it for convenience of feasibility, as walking is rarely available for such trips (6.59%), and the choice seems to be mainly between public transportation and own car. Most hospital visitors report low income (50.55%), and only a few report no income (6.59%), further supporting the argument. Another interesting fact is that most car travelers report low income (48.39%). Although the present study does not directly analyze the impact of car ownership as an explanatory variable, a case can be made for its higher importance in choosing to drive compared to income level.

The coefficients estimated on MNL are not statistically significant for riding with someone. Nevertheless, the same pattern as in own car is observed in LCM2, with a strong positive association between riding with someone and enjoying a high income and the opposite being true for low income. In LCM1, although negative, both coefficients have a small magnitude, specifically in comparison to other alternatives, suggesting some degree of independence from income on this alternative. These findings are mostly in line with own car, although not as conclusive, and applicable to individuals with the same preference patterns.

Possessing any form of income increases ride-hailing's utility severely, as per LCM1. A positive coefficient for low-income individuals in LCM2 is the only remaining significant effect. Overall, the likelihood of choosing ride-hailing increases with income, particularly for low levels. Conversely, the results for a more traditional service like taxi are inconclusive.

According to MNL, high income has a positive impact on cycling utility. This effect is, to some extent, also observable on LCM, with the same scenario also applicable to low-income for both classes—though with small relative magnitudes.

Age, sex and household composition

All the results for the excluded factors are relative to MNL estimations. This fact also applies to the excluded factors described at the end of Sections 4.2.3 and 4.2.4.

Although BS1, BS2, and HD include age, the estimated coefficients for this factor were consistently non-significant for every alternative, with the only exception being taxi. As explained,

any conclusion regarding this alternative is most likely invalid. This study concludes that age is likely irrelevant to mode choice in the present context. Such might not hold for a model that does not explicitly account for the effect of income level, as both factors are seemingly correlated.

The case of sex is more alarming than the one for age: although the factor complied with the rules for inclusion in BS1, it was shortly found to possess only non-significant coefficients, resulting in its exclusion from BS2. Again, this is a likely irrelevant factor for mode choice in a sample with an excellent balance of male-female answers (close to 50-50).

The coefficients of household composition estimated on the most extensive set of factors (Full) showed statistical significance for most alternatives. When applying any of the significance-driven selection strategies, this significance disappears. Although a case can still be made for its importance, considering it of minimal relevance for mode choice in this setting is also reasonable. Removing convenience and/or trip purpose might weaken its explanatory role in mode choice, suggesting potential interactions that require further investigation.

4.2.3 Objective factors

Summary of main insights: (1) More regular trips arguably favor metro over bus, with a notable preference for driving for highly regular trips. (2) Public transportation—metro in particular—and riding with someone are associated with more multi-modal journeys; own car links more to direct trips without additional modes. (3) Walking is interestingly more likely for reasonably longer trips. (4) Private modes and ride-hailing are more likely preferred at higher costs, possibly due to perceived flexibility or necessity. (5 & 6) Individuals are generally more sensitive to additional waiting/walking times than additional trip duration, i.e., time inside the vehicle; metro generally attracts individuals with more tolerance to additional walking, while own car is less attractive when additional walking is required. Trip purpose requires additional research, as it showed a lack of significance. Whether a trip is regular or occasional is likely irrelevant to mode choice.

Trip regularity

To best comprehend the results of this factor, the reader must have in mind that, before imputation, 86.68% of regular trips are associated with the highest possible regularity category (3+ times per week).

Although the MNL estimates positive coefficients for trip regularity for public transportation, the same does not apply to LCM. Regardless, the estimated coefficient for metro has a more substantial magnitude, which must be considered. In the more sophisticated approach, bus is associated with a significant negative coefficient for both classes. Metro, in turn, also possesses a positive coefficient for both classes—this coefficient is, nonetheless, irrelevant in LCM2 for lack of magnitude and statistical significance. The present study finds that individuals are most likely to opt for metro over bus when choosing a public transportation alternative for a highly regular trip. The data support this: metro is the alternative whose users report the highest trip regularity per week, on average (~3.80 times per week); for users reporting the highest regularity possible,

metro is also the most popular alternative, despite being chosen fewer times than bus across all trips.

The interpretation of this factor for walking depends highly on the approach used. While presenting a positive coefficient in MNL, this factor is significantly close to zero in LCM1 and negative in LCM2, but still with a relatively small magnitude. Using the MNL, the obvious suggestion is that more regular trips increase the likelihood of choosing to walk. However, LCM does not support this. If anything, there is a subtle tendency to walk less for more regular trips, potentially favoring modes such as metro or own car.

Given the positive coefficient registered for own car in all the studied instances, including a coefficient with an extremely high magnitude on LCM2, this study concludes that the higher the regularity of the trip, the higher the likelihood of choosing own car. In comparison to metro, this likelihood is also higher, according to LCM. Consequently, this study further concludes that own car is a more likely alternative than public transportation for individuals taking a specific trip more regularly. This finding is not directly observable since own car appears to be used less regularly than average (3.03 vs. 3.61 times per week) and is used almost equivalently for occasional and regular trips. However, one must recall that the inference derived from the estimated coefficients makes sense only when all alternatives are available—which is an impossible scenario for individuals who do not own a car—but also that higher regularity already implies that the trip is not occasional.

Discrepancies observed in the sign of the coefficient for riding with someone suggest, once again, the heterogeneity associated with the population—this time, in the way they yield more or less utility from more or less regular trips. Nevertheless, since a negative sign appears in LCM2, most individuals are less likely to ride with someone for more regular trips. Such a finding is also applicable in comparison to all the aforementioned modes. There is also a high degree of heterogeneity in the case of ride-hailing. While a positive sign in MNL and a negative sign in LCM1 may suggest inconclusiveness, an extremely strong positive coefficient is estimated in LCM2, suggesting that most individuals prefer ride-hailing for more regular trips. Although this is counterintuitive at first glance, it may support the theory that individuals use this mode in specific contexts because of the lack of other feasible alternatives for trips the individual needs to take. It is in LCM2 that ride-hailing scores the most substantial negative coefficient for perceiving the chosen alternative as first choice—and by a long margin. Section 4.2.4 contains further recommended reading on this matter.

For bicycle/scooter, the estimated coefficient in LCM1 is positive, yet almost zero. By looking at the other results, it is observed that the overall tendency is a negative impact of more regular trips on utility, with negative coefficients registered in MNL and LCM2—where this effect is particularly strong. With the available sample, the present study finds that individuals tend to avoid bicycle/scooter for more regular trips. As a matter of fact, this mode is the alternative associated with lower trip regularity on average (2.33 times per week), except for taxi. The effect of ownership/availability, although not measured explicitly in the model, can be observed and also seems to support the conclusions. Individuals chose it only 3 times out of the 36 times it was

available. For taxi, the results are mainly inconclusive and/or show independence from this factor.

Number of modes

According to MNL, the number of modes associated with a traveling journey represents a positive utility for all alternatives. Only the coefficient for taxi is non-significant, and, in terms of magnitude, there are no considerable differences, except for own car, which exhibits a noticeably weaker coefficient compared to the remaining modes. This parameter, by itself, means that a higher number of modes links with a lower probability of choosing own car in comparison to any alternative. It also suggests that cars are associated with fewer multi-modal journeys, which is directly observable in the data, as no individual who chose own car reported having to resort to an additional alternative (besides walking) to complete their journey. These findings suggest that own car is a more convenient and flexible alternative for more direct traveling.

The comparison between own car and public transportation—and public transportation within itself—becomes even more interesting when evaluating LCM. LCM1 shows a relatively small positive coefficient for own car, pointing to potential independence from number of modes. However, the only significant negative coefficient in LCM2 is found precisely for this alternative, suggesting that previous conclusions apply to most individuals. In contrast, in LCM1, bus presents a strong positive coefficient and metro a negative one; in LCM2, bus has a non-significant coefficient close to zero, and metro, unsurprisingly, displays a severely strong positive coefficient. Such ultimately means that bus is likely associated with more multi-modal journeys for a portion of the population than metro. Nevertheless, the opposite is observed for a larger portion of the population, accompanied by a substantially stronger effect. In fact, on average, metro is the mode associated with more multi-modal journeys (1.40 modes + walking). Bus comes fourth (1.19 modes + walking).

According to LCM, riding with someone is also associated with more multi-modal trips, particularly for LCM2. This finding extends the argument presented in Section 4.2.2 for the possibility that this mode is typically more used with others to take a given journey, namely with public transportation services.

Ride-hailing results remain consistent for all scenarios, indicating that this mode is more likely to be chosen as part of a journey when its number of modes is higher. Such is primarily true according to LCM1. Bicycle/scooter presents a similar pattern but with a milder effect. Whether or not these alternatives can be used as a last-mile solution when walking is not feasible requires further investigation. For taxi, the results are mainly inconclusive and/or show independence from this factor.

Trip duration

In MNL, trip duration is significantly perceived as a positive utility for the most popular alternatives, i.e., bus, metro, walk, and own car. Among these, walk presents the most substantial coefficient, which might seem counterintuitive and calls for a more nuanced analysis.

The sign is inverted only for bus in LCM1. It is, nonetheless, close to zero for own car and particularly close to zero for metro, implying individuals with such characteristics as defined by LCM1 are almost non-sensitive to trip duration. In LCM2, the situation changes: bus presents a positive coefficient, metro a robust negative coefficient, and own car a negative coefficient. Hence, when choosing between a public transportation alternative and own car for longer trips, a portion of the population will tend to favor own car and then metro, avoiding bus. For a larger portion of the population, however, bus is far more likely to be chosen, while metro emerges as the least likely alternative. This result might suggest that most individuals using metro and own car might be more short on time and do not value the time inside the vehicle. Traveling by metro might also be more associated with a higher lack of comfort (traveling standing and/or with crowded space). Time spent in traffic can potentially influence the trip duration for those who drive.

Walking displays no discrepancies throughout the results. Although walking trips are typically the shortest, except for taxi, individuals seem to enjoy the time they spend walking and might value its benefits—e.g., health or relaxation. Another important consideration is the lack of substantial variation in trip/journey duration across the sample. Actually, the fact that the most extended option available as an answer for these questions was *15+ minutes* (approximated to 17.5 minutes) places the present study within the context of short-distance travel, where walking is considerably more feasible. Trips are arguably not sufficiently long enough to discourage individuals from walking. It is also plausible that the relationship between utility and trip duration is not one of strict linearity as it was modeled: individuals may have specific perception thresholds when evaluating this factor, as demonstrated in Cantillo et al. (2006), and/or possess unique interactions that can only be captured by different model identifications. Similar conclusions are applied to bicycle/scooter by observation of LCM, extending the argument that these active modes are typically more enjoyable and yield unique benefits.

Results for riding with someone display inconclusiveness in MNL and contrasting preferences in LCM, specifically registering a negative coefficient in LCM2. These contrasting preferences are also captured for ride-hailing, this time with an especially weak positive coefficient on LCM2. If anything, the impact on utility that most prevail as trip duration increases is the one from LCM2 for people riding with someone since more individuals are expected to belong to this class, and the one from LCM1 for ride-hailing, as LCM2 exhibits lower sensitivity, while LCM1 demonstrates a significantly stronger effect. Thus, as trip duration increases, individuals are less likely to ride with someone or choose to travel with a ride-hailing platform.

According to LCM, individuals do not consider trip duration when choosing a taxi, contrary to MNL, where the coefficient points to lower choice likelihood with increased duration. Again, the lack of support makes extrapolation impossible.

Cost

All the estimated coefficients for cost in MNL are positive. The result is in clear opposition to what is expected according to economic theory since, as in the case of trip duration, individuals typically perceive these factors as a *disutility*. Additionally, there is minimal variation in the

estimated values, except for the coefficients of taxi and bicycle/scooter, which are not significant, meaning that the likelihood of choosing between public transportation (bus and metro), car options (own car and riding with someone), and ride-hailing is not substantially cost-dependent. Despite that, at the exact cost, there is a slight lean towards avoiding public transportation, though, again, the differences are negligible. Train (2009g) provides an interesting explanation as to whether such effect is captured: using data from revealed preferences, as in the present study, factors that are the most important to consumers typically exhibit low variation due to the natural forces of market equilibrium. Nevertheless, the situation changes when using LCM, providing a better scenario illustrating differences between modes.

Starting with public transportation, a negative sign for both options in LCM1 aligns with what is typical from this factor—with bus presenting a more substantial magnitude coefficient. In LCM2, metro is the only alternative presenting a negative sign. Consequently, this study concludes that, for a portion of the population, individuals yield lower utility from higher costs for both public transportation options, diminishing the likelihood of choosing them compared to, e.g., own car. When evaluating this mode type, these can be considered the most cost-sensitive individuals—in a negative sense. Moreover, for this segment of the population, when faced with the two options, individuals are more likely to choose metro for the same ticket price, i.e., although perceiving cost as a *disutility*, they are slightly more willing to pay to travel with the metro. In LCM2, the opposite is applicable, meaning that most individuals will avoid metro at higher costs. The perception of cost associated with bus captured by LCM2 arguably reflects that most individuals find this mode to present a higher quality of service than metro. Considering the subscription effect, the result for this class can also be cross-examined. Since individuals in this class are more likely to choose metro when owning a subscription, they might not be as willing to pay an additional amount besides the fixed cost that the subscription might already represent. Conversely, if individuals in this class do not own any subscriptions, it can be concluded that they are more likely to prefer paying to travel by bus.

For own car, the coefficient is still positive for both classes, in stark contrast to metro. It also presents the most substantial positive coefficient in LCM1, with the same happening in LCM2, excluding ride service options (ride-hailing and taxi). Therefore, the present study concludes that, globally, own car is more likely to be chosen for more costly trips. Another way to interpret these results is that for prices showing slight variation, own car is highly more likely to be chosen than public transportation. Again, the effect of car availability/ownership is expected to play a critical role in such a decision-making scenario. Indeed, only 118 out of 687 individuals report having own car and both public transportation options available. In these instances, own car is the most chosen option, with an outstanding 44.07% share, higher than the combined share of bus and metro. Although counterintuitive at first sight, the results are comprehensible. Though not genuinely wanting to pay more, individuals may link car driving with desirable features like higher flexibility, safety, comfort, and/or a sense of exclusivity. This alternative can also be associated with a personal lifestyle and/or perceived as status-enhancing by many individuals. The costs are not significantly high enough to deter driving—as trip duration might not be long enough to

discourage walking. Likewise, the way the study measured cost heavily relies on travelers since it was reported by themselves and not estimated via ticket price or traveling history data, for example. In this context, an important remark is that individuals choosing own car seem to have a higher difficulty gauging the actual cost of a trip. 35.48% is the percentage of those who do not know how much they paid for the trip, which is the most common answer for these individuals. A potential underestimation of actual costs is possibly also responsible for increasing the tendency to choose this mode over other alternatives.

In the case of riding with someone and ride-hailing, the coefficient is also positive for both classes. Riding with someone's coefficients are always smaller than own car's, particularly in LCM1, where this coefficient has a small relative magnitude. Ride-hailing, in its turn, presents the most substantial positive coefficient in LCM2. The present study concludes that, globally, individuals are more willing to pay to drive their car than to ride with someone, though the situation is quite atypical (31 out of 687 times). It is also possible to state that, for most individuals, ride-hailing is preferable at higher costs. Again, this does not necessarily mean that they prefer to pay more. It reinforces the idea that individuals choose the alternative because of other alternatives' lack of feasibility or flexibility since the effect is captured in LCM2 (associated with lower perceived and systematic preferences regarding ride-hailing).

Given the presented coefficients for bicycle/scooter, individuals do not seem substantially cost-sensitive concerning this alternative. It might not also be appropriate to judge cost concerning this alternative since most bicycle trips arguably represent no direct cost, and it is not known if the reported cost is related to a scooter or bicycle. For taxi, similar conclusions could be derived from the available information for LCM1, as it presents a virtually null coefficient. These conclusions are speculative for the aforementioned reasons and allude again to the lack of data for these alternatives.

Walking and waiting times

It is crucial to comprehend the preferences of walking and waiting times simultaneously with trip duration to understand better how the total trip travel time influences decision-making. Regardless of the sign presented by the coefficients, individuals seem consistently more impacted by waiting time than by walking time and trip duration, according to MNL. The only comparison including the three factors, given their applicability, is for public transportation, where such a conclusion is evident: waiting time's coefficient is, on average, $\sim 3.0x$ stronger than the one for walking, and $\sim 3.7x$ stronger in comparison to trip duration's.

In LCM, though, a higher equilibrium is registered between walking and waiting time, with an edge to walking time; trip duration still represents a lower impact on utility for a minute increase/decrease. Using public transportation, as well, to illustrate this situation: in LCM1, waiting time's coefficient is, on average, $\sim 1.1x$ weaker than the one for walking, and $\sim 9.0x$ stronger in comparison to trip duration, where the effect is exacerbated by the low trip duration coefficient for metro; in LCM2, waiting time's coefficient is, on average, $\sim 1.7x$ weaker than the one for walking, and $\sim 2.2x$ stronger in comparison to trip duration's. Consequently, the present study finds that,

for public transportation, individuals generally see a higher impact on utility from an additional time unit of walking/waiting time compared to trip duration. However, there is a significant caveat to these results: the coefficients show positive signs in every instance. Similarly to cost, this does not necessarily mean that individuals want to walk and/or wait longer to use public transportation. It may, conversely, mean that these modes typically oblige people to wait/walk for higher amounts of time in comparison to other alternatives but that they are still willing to wait and/or walk, at least within what is reasonable. The availability effect can also play an important role in making such a hypothesis accurate, as 234 out of 687 individuals reportedly only had at least one of the public transportation alternatives to take their trip. Therefore, a proper comparison is to evaluate how these factors influence the choice between only metro and bus (registered 116 times). According to MNL, the differences between these alternatives are not substantial for both factors; hence, no conclusions can be inferred. In LCM1, metro is more sensitive to walking time and bus to waiting time. In this class, individuals are likely more willing to walk to catch the metro but less willing to wait for it. In LCM2, to the contrary, metro is more sensitive to variations for both factors—though the difference to bus is not as severe as in LCM1—indicating that these individuals yield marginally higher utility when waiting and/or walking for the same time as they would do to take the bus, thus being more willing to wait and/or walk to travel by metro. Overall, although the same might not be as straightforwardly applicable for waiting time, individuals are consistently more willing to walk to catch the metro. Alongside the systematic preference effect already described, this might help explain the slightly revealed tendency for metro for the given scenario (66 out of the 116 times) since, on average, walking times are practically the same for the bus and (5.35 minutes for metro vs. 5.38 minutes for bus).

Except for bicycle/scooter, the remaining applicable alternatives show non-significant results in MNL for walking time. When resorting to LCM, LCM1 shows a negative coefficient for own car and a marginally positive coefficient for riding with someone. In LCM2, the sign for own car is inverted, though its magnitude is still considerably lower compared to public transportation; riding with someone's coefficient, in turn, is approximately two times weaker than own car's. These results suggest that, compared to public transportation, both car alternatives are significantly less likely to be chosen if such a decision inherently means a higher additional walking time. LCM2 exhibits lower sensitivity, while LCM1 demonstrates a considerably more substantial effect. In LCM2, the effect is diluted so that riding with someone becomes the less attractive alternative for the same walking time. In this portion of the population, individuals seem more willing to walk longer if it enables access to a more systematically preferred alternative, as in the case of own car. In essence, walking time for car alternatives is arguably perceived as more avoidable or an inconvenience stemming from the effect of parking availability, contrasting with the more accepted and structural walking component of public transportation. For own car in particular, longer walking times might severely disrupt the convenience and flexibility narrative associated with driving a private vehicle, making public transportation relatively more attractive. The effect of accessibility and other built environment factors might play a role in the present context, although, once more, further investigation with a more targeted survey would be needed. Bicy-

cle/scooter is the least attractive alternative when accounting for additional walking time. This mode is arguably a more flexible alternative than own car in terms of parking, which strengthens the previous argument. It may also suggest that individuals anticipate walking considerably less when using such alternatives. The available sample reinforces the idea: all three individuals report that using bicycle/scooter adds the least possible amount of walking time to their trip, i.e., *less than 1 minute*.

In MNL, both ride service alternatives present similar results to public transportation regarding waiting time. When analyzing LCM, the most substantial difference between ride-hailing and public transportation is its weaker coefficient, even if still positive. In line with previous explanations, most individuals are slightly less likely to tolerate additional waiting time to take a trip using a ride-hailing platform when public transportation alternatives are available. On average, ride-hailing users do not, in fact, need to wait as long as public transportation users. Taxi presents the lowest coefficients for waiting time. Therefore, the results indicate lower sensitivity to waiting time, hence a lower likelihood of being chosen if the wait is significant. This inference is merely exploratory, nevertheless.

Trip purpose and trip character

Trip purpose was removed after the estimation with all factors with MNL for reasons cited in Section 3.3. Nevertheless, in line with those reasons, evaluating whether each trip's purpose could impact mode choice in the present context is still important. *Hospital visits* significantly impact only riding with someone and own car. *Leisure* impacts almost every alternative, though only walking at a highly significant level ($p < 0.01$). Both facts might aid the hypotheses presented for income level in Section 4.2.2. *Work* trips were, surprisingly, deemed not significant for any mode. This result might mean no significant differences in mode choice for more mandatory trips like commutes to *work* and *study*, fomenting further investigation. *Shopping* and *other* purposes are examples of generally relevant purposes alongside *leisure*, arguably constituting less mandatory trips. Overall, the cons outweigh the pros of specifying these variables in the models, as the trip purpose factor, in its present configuration (refined in 8 different categories and measured along 8 different alternatives), seems of limited impact on mode choice. In both estimations using the sets of factors derived from *simple backward selection* (BS1 and BS2), whether a trip is regular or occasional is only statistically significant for walking trips.

4.2.4 Subjective factors

Summary of main insights: (1) Lower satisfaction levels for a trip are generally expected from public transportation users, in stark contrast to pedestrians. (2) Ride-hailing is avoided according to individual perception, aligning with systematic preferences. Convenience showed no impact across all alternatives, while feeling of control, though excluded from the presented models, is plausibly relevant for only a few modes. These findings predominantly align with observed data patterns and underscore the nuanced role of psychological factors in mode choice behavior.

Satisfaction

Overall, high levels of stated satisfaction for a given trip are consistently associated with a lower likelihood that the trip was made using any of the public transportation alternatives. This finding aligns with the data, which shows that public transportation users display below-average satisfaction levels despite overall satisfaction across modes.

Walking and own car, although exhibiting a negative coefficient for this factor in MNL, display a completely inverted scenario in LCM for both classes—with walking possessing greater coefficient magnitude in both instances. Bicycle/scooter, in turn, has a significant positive coefficient in all possible scenarios. Again, these results follow the data, as car owners and pedestrians are among the individuals who report the highest satisfaction levels, and cyclists report the highest experience-related scores, including satisfaction. The literature also supports such, specifically for the active modes.

For ride-hailing, MNL shows inconclusive results, and LCM, alternatively, proposes a significant heterogeneity discrepancy, with coefficients in distinct classes possessing opposite signs. The sharp negative perception of this mode reported by individuals associated with LCM2 and the alternative's systematic *dispreference* might explain this contrast. Since individuals potentially choose it as a last resort, the stated satisfaction immediately after the trip might tend to be poorer for them.

Results for riding with someone are either inconclusive or show independence of reported satisfaction, except for LCM2, where this factor represents a positive utility for the alternative. Finally, for taxi, the available information leans towards similar conclusions as in bicycle/scooter, although only for a portion of the population, but still as speculative.

Perception as first choice

If an individual agrees that the chosen mode used for a trip was their first choice for that given trip, the likelihood of the individual having chosen to walk is significantly higher according to any of the studied scenarios. The same is valid for bicycle/scooter, though only supported by LCM, for lack of statistical significance in MNL. Concerning own car, the findings point to independence or inconclusiveness, except LCM1, where the estimated coefficient is positive, indicating the same conclusions as for walking, at least for a portion of the population. All conclusions align with the data, as these three alternatives score the highest for this factor. Furthermore, this is the factor that arguably best distinguishes the alternatives based on reported mode experience.

Although MNL points to inconclusive results for public transportation, the bus' coefficient is positive for both classes in LCM, and the metro's coefficient, although negative in LCM1, shows a strong positive utility impact in LCM2, evidencing congruence with systematic preferences for individuals with the characteristics illustrated by this class. The effect is a specific case of perceptions aligning with regular behavior. Again, in congruence with systematic preferences, the coefficient for ride-hailing is always negative, particularly with an extremely strong magnitude in LCM2. Another compelling observation from the data is that ride-hailing is the alternative whose

users affirm that they are least willing to travel again, which can be somewhat interrelated with the findings regarding this mode. The results of riding with someone and taxi are mainly divisive and/or inconclusive.

Convenience and feeling of control

Convenience was the only factor showing no statistical significance for every alternative after the estimation employing all factors considered for the modeling phase. Although this factor is highly likely not to impact mode choice in the present research context, it is recommended to include it in other studies, mainly if measured in a distinct manner, e.g., as a proxy combining other factors or measurements.

In the estimation using BS2 (see Table 3.9), feeling of control was deemed significant for the three most popular alternatives, plus ride-hailing. Although the models presented in this study do not include this factor—for non-conformity with the *Backward selection with threshold* strategy after the estimation on the Full set (see Table 3.9)—, it is plausible that it still holds some relevance for mode choice in the present context. The coefficients of public transportation alternatives and ride-hailing are negative in the mentioned estimation, suggesting that these modes are less likely to be used by individuals who want to feel in control during the trip. The opposite is applicable for walking. Although these conclusions seem to be directly supported by the data—public transportation and ride-hailing are associated with the worst scores for feeling of control, while walking is the second best performer, after bicycle/scooter—it may be problematic to accept them since it is not guaranteed to yield similar results on the presented models by simply adding this factor to their respective specifications.

4.2.5 Willingness to pay

In discrete choice modeling, evaluating the *substitution pattern* inherent to any pair of attributes is possible. When a model accounts explicitly for a cost variable, it allows for one specific case analysis in the context of such an evaluation. It corresponds to analyzing the *substitution pattern* by using cost as a reference attribute, i.e., analyzing hypothetical amounts individuals would be willing to pay to see specific modifications in a second attribute. This is known as *willingness to pay* (WTP) (Train, 2009a). A special case of this measure occurs when the *substitution pattern* is evaluated for a time-cost pair, with, in such a case, WTP representing what is also named value of time (VOT) or value of travel-time savings (VTTS) (Hess et al., 2005).

Computing such economic measures is straightforward, particularly in the case of fixed taste coefficients, as in MNL. For practical issues and easier comprehension of the results, the final estimation using such a model was chosen for analysis. It is prudent to keep in mind that the mentioned analysis might not necessarily reflect previous conclusions as a whole, given that the results of LCM heavily influenced them. Nonetheless, observing the cost-benefit scenario for dimensions individuals perceive as crucial for mobility is still insightful.

Using the typical linear-in-variables utility specification, as in the present study, WTP can be derived as the negative ratio of the analyzed coefficient and the cost coefficient:

$$WTP = -\frac{\beta_k}{\beta_{Cost}} \quad (4.1)$$

As seen previously, all estimated cost coefficients in MNL are positive. Although this does not invalidate the WTP analysis, per se, it requires a more nuanced interpretation. The WTP formula presented in Eq. 4.1 leads to a negative value if both coefficients are positive. Such is rational, as cost must decrease to compensate for the increase of attribute k . In these cases, WTP is typically denoted as willingness to accept (WTA), meaning that the individual would require compensation or incentive for the modification to take place. WTP can be evaluated as the reflection of an attribute's unit increase or decrease depending on what is more intuitive and logical in terms of human behavior. Eq. 4.1 is formulated concerning the scenario of an increase. Likewise, using the example of number of modes, it might be more appealing to evaluate how much the individual is willing to pay to reduce the number of modes associated with a journey than to be compensated by an increase. In this alternative situation, the correction is made by removing the negative sign in the considered ratio:

$$WTP_{decrease} = \frac{\beta_k}{\beta_{Cost}} \quad (4.2)$$

Among the alternatives studied in the present research, walking is excluded from the analysis since it has no inherent applicable cost. Likewise, bicycle/scooter and taxi were excluded because of their low support. The computation of the measure was deemed viable only when both estimated coefficients—for cost and the studied factor—were statistically significant at a 5% significance level. Moreover, the viability depends on whether the trade-off is applicable in the real world: e.g., it does not make sense to study how much individuals are willing to pay to increase their income level. For this reason, the analysis did not include income level, owning a subscription (cost-defined, *a priori*), and cost itself.

Two known limitations of the presented analysis are the assumption of linearity regardless of the factor's values, plus the fact that the effects studied are deterministic, i.e., do not account for the error term (Hess et al., 2005). For example, the assumption of linearity might oversimplify the preferences associated with factors like trip regularity or waiting time, where an effect of diminishing marginal utilities could come into play. Although it does not reflect exact human behavior, the approximation can still be reasonably accurate for relatively small variations of the studied factors.

WTP for a unit increase

Table 4.5 illustrates the results from the WTP computation, first unveiling negative values for trip regularity. Such values do not necessarily mean that individuals dislike traveling more regularly; they expect a cost reduction to accept even more regular trips. Such might be explained by a

Table 4.5: Willingness to pay in euros (€) for the parameter estimation results from MNL using set of factors TT2.

Factor	Bus (<i>N</i> = 240)	Metro (<i>N</i> = 220)	Own Car (<i>N</i> = 93)	Riding with someone (<i>N</i> = 16)	Ride-hailing (<i>N</i> = 15)
<i>WTP for a unit increase</i>					
Trip regularity	-0.12 a	-0.20 a	-0.18 a	-0.21 a	-0.15
Perception as first choice					0.14 a
Satisfaction	0.32	0.34	0.29	a	
<i>WTP for a unit decrease</i>					
Number of modes	2.60	2.89	1.22	2.62	2.74
Trip duration	0.06	0.04	0.04	a	0.04
Walking time	0.06	0.07	a	a	b
Waiting time	0.20	0.18	b	a	0.16

^a Either cost or analyzed factor's coefficient not statistically significant at a 5% significance level.

^b Not applicable for the respective alternative.

reluctance to increase the regularity of the trip if it is already reasonably aligned with personal necessities. As an illustration, for metro users, who typically report the highest regularity on average, the 20-cent compensation for traveling with this mode once more per week is arguably relevant for the current prices since the average metro cost is 2.21€. For the average individual, accepting this incentive and adding a metro trip to their weekly routine would translate into 9.65€ spending for 4.8 trips per week, compared to the current 8.40€ for 3.8 trips. This result translates to a ~15% cost increase for ~26% more service usage. Again, whether or not individuals would be interested in this trade-off requires further investigation.

Individuals are willing to pay more to change how they perceive ride-hailing as their first choice. This finding might link to the desire for service enhancements or higher availability. For the given alternative and factor, the average score is 2.93 out of 5, indicating slight *disagreement* with considering it a first choice to travel. An additional 1.07 points would be needed to reflect an overall *agreement* (equivalent to 4 out of 5 on a 5-point Likert scale), on average, among individuals. This increase is associated with 0.15€ higher costs, amounting to a ~3% average cost increase for a markedly improved mode perception.

In turn, the results for satisfaction reveal a higher individual sensitivity to this factor compared to perception as first choice, even if evaluated for different modes. Bus (3.95), metro (4.04), and own car (4.27) already show moderate levels of satisfaction. In a scenario where the population choosing these modes is divided between *agreeing* and *fully agreeing* with being satisfied (4.5 out of 5), additional costs would be 0.18€, 0.16€, and 0.07€, respectively. Public transportation shows a more significant potential for satisfaction improvements. Similarly to the results of perception as first choice, these values are possibly also linked with overall service quality enhancements, with the willingness to pay for such enhancements reflected in these values.

WTP for a unit decrease

The analysis for number of modes results in the most substantial WTP values, as shown in Table 4.5. The straightforward interpretation for such high figures is that, for this factor, WTP represents what individuals are willing to pay to avoid a more complex journey, i.e., the amount they would otherwise pay for an additional trip in a whole journey. Individuals are particularly more willing to pay for more direct public transportation trips. Regarding these alternatives, metro users would be willing to pay $\sim 131\%$ more than what they spend on average for a trip, as long as it would save them an additional trip. The feasibility of such a scenario is highly dependent on the quality improvements in terms of public transportation infrastructure and available routes. Though still reflective of a preference for reduced transfers, the percentage for bus is substantially lower ($\sim 79\%$).

Finally, individuals display higher sensitivity to waiting time according to MNL, as mentioned in Section 4.2.3. The considerably higher willingness to pay to reduce this time, in comparison, e.g., to the time inside the vehicle, reflects this finding. Since individuals spend noticeably more time inside the vehicle, the impact of one added/subtracted minute is naturally less perceived and more residual. Metro users wait 4.35 minutes, on average. According to the results, they would spend an additional 0.39€ to see this time regularly cut in half, undermining the arguably irritating effect of the factor on individuals. Overall, WTP for these factors showcases the importance of on-demand availability and frequency of metro, but with extendable applicability to bus and ride-hailing. Furthermore, at least with MNL, the results might also suggest that the infrastructure is already accessible within reason.

Chapter 5

Conclusions and future research

This study investigated the underlying mobility patterns at the Asprela university campus, resorting to discrete choice models to examine the impact of various socio-demographic, objective, and subjective factors. These models harnessed a sample collected via an RP survey that included eight different mode alternatives, with bus, metro, walking, and own car among the most frequently chosen. It contributes to a theoretical and practical understanding of discrete choice in urban mobility. In practice, targeted interventions can be derived to promote a more sustainable and efficient commute but are not explicitly promoted. Incorporating the computation of different consumer surplus scenarios in future developments can allow for more robust recommendations regarding policy implications. For instance, targeted subsidies for public transportation or investments in infrastructure to enhance accessibility could be informed by the suggested methodology.

Notwithstanding inherent limitations derived from the well-known IIA property, the traditional MNL, in addition to being used as a reference model, also proved its powers in discrete choice analysis. The straightforward WTP analysis it yielded underscores such advantages. Nonetheless, its assumption of homogeneity across all individuals inside the studied population restricts more nuanced insights into the travelers' highly complex decision-making process. With the proposed LCM, the limitations can be partially dismantled by simply allowing the population to be segmented into two groups with distinct preference patterns. Combining the estimation results of both models made a more comprehensive assessment possible, as the values often indicated slight discrepancies potentially inherent to the population's heterogeneity. These discrepancies are not dissatisfactory. Instead, they should be valued, for they potentiate a more in-depth discussion of results instead of a one-sided argument. The ability to capture individual heterogeneity can be further enhanced in even more sophisticated models using simulation-based estimation, such as MMNL.

Conversely, in the present context, nested extensions of MNL showcased some of the limitations mentioned in the literature. A particular case was the optimization problem that arose when performing the MLE for these models. Such methodological findings have value and should foment additional research on applying nested models to scenarios with several alternatives, as in this case. Nonetheless, the improved validation results yielded by the traditional NL and the flex-

ibility in capturing more realistic patterns should be enough stimuli to overcome the mentioned limitations.

Other methodological findings derived from the implemented validation process are worth mentioning. MICE unsurprisingly emerges as a practical tool for robust missing data imputation results, which can be highly useful to complement surveys that allow for *I don't know / I prefer not to answer* options. Using more straightforward, less over-parametrized models can be of higher use if their goal is more geared towards forecasting travel demand. A puzzling effect that would require more research is the experimentation with even more diverse factor selection criteria, e.g., stepwise regression. Inside the strategies used, slight changes could also be carried out, like keeping factors with significant coefficients for two out of the four most frequently chosen modes or introducing a higher level of restrictiveness by considering a 1% significance level.

In line with Occam's razor, the advantage of simplicity is applicable when comparing sets of factors with different dimensions and choosing among different model types. For instance, it was found that LCM, for its higher complexity, is more prone to yield a better fit, which can, sometimes, be exacerbated and translated into an overfit and/or a lesser performance in a prediction context. Nevertheless, the final versions of the MNL and LCM provide highly reasonable results for these aggregate outcomes. Additionally, favoring out-of-sample validation and proper validation metrics instead of relying solely on goodness-of-fit statistics extracted from in-sample validation sets the ground for these findings while promoting more realistic and optimal models. This relatively simple approach allows exploring more complex and computationally challenging solutions involving cross-validation and/or coefficient estimation using multiple datasets.

Other fascinating options, as cited in Section 2.2, could be applicable and are related to: the integration of advanced machine learning techniques, leveraging the pros associated with more traditional models with more innovative ones; probabilistic choice set generation as an intermediate step, given the apparent correlation of some availability variables with measured factors, as illustrated in Section 3.1. In terms of studied factors, further modeling refinements can also include car availability as an explanatory variable. A case can also be made for household composition, trip purpose, and feeling of control. Likewise, future surveys on urban mode choice are encouraged to inquire about built environment-related factors.

The key conclusions are as follows for the remaining (groups of) factors, i.e., the modeled ones. Systematic preferences deeply impact mode choice: own car, metro, and walking—in descending order—are inherently more attractive than bus for reasons not captured by observed factors. Inside socio-demographic factors: owning a subscription stimulates public transportation use while strongly discouraging walking trips; higher-income individuals are likelier to walk and drive than to choose public transportation; age and sex, typically noteworthy in other settings, are likely irrelevant in the studied context. Objective factors seem to have the most decisive influence on mode choice in Asprela: individuals are more likely to favor metro over bus for highly regular trips; public transportation, especially metro, is expected to play a crucial role in multi-modality; walking is surprisingly more likely chosen for reasonably longer trips; private modes, e.g., own car and ride-hailing, are more likely chosen at higher costs; additional waiting/walking times have

a higher impact than additional trip duration, with own car being a less attractive alternative if additional walking is required; a trip being occasional or regular is not relevant. Lastly, subjective factors require a more cautious interpretation: lower satisfaction is generally expected from public transportation, in contrast to walking; ride-hailing is substantially avoided according to individual perceptions, which aligns with systematic preferences; convenience does not impact any alternative, though its modeling is highly prudent in any case. In the context of WTP, price discounts are needed to incentivize more regular trips. Moreover, small cost increases can improve moderate satisfaction levels, with individuals potentially willing to invest in improved service quality. There is also a strong individual willingness to take more direct trips while being highly sensitive to waiting time.

Regarding the specificity of the results from this setting for broader inference on urban mobility, it is advisable to keep implementing these models in distinct contexts and geographical areas, particularly—if possible—performing the necessary methodological alignments for the feasibility of external validation. Such would significantly increase the confidence in the said models' generalizability and results, leading to even more conclusive results.

Finally, this research underscores the importance of more tailored, context-specific solutions for mode choice and the severe complexity inherent to discrete choice analysis, which, contrary to expectations and intuition, is often more challenging than analyzing continuous goods. By comprehending the evidenced factors, all stakeholders involved with transportation are invited to address the future challenges posed by urban mobility, paving the road for an evolved, state-of-the-art transportation ecosystem.

Bibliography

- Abou-Zeid, M., Witter, R., Bierlaire, M., Kaufmann, V., and Ben-Akiva, M. (2012). Happiness and travel mode switching: Findings from a swiss public transportation experiment. *Transport Policy*, 19:93–104.
- Acker, V. V., Kessels, R., Cuervo, D. P., Lannoo, S., and Witlox, F. (2020). Preferences for long-distance coach transport: Evidence from a discrete choice experiment. *Transportation Research Part A: Policy and Practice*, 132:759–779.
- Ahmad, S. and Puppim de Oliveira, J. A. (2016). Determinants of urban mobility in India: Lessons for promoting sustainable and inclusive urban transportation in developing countries. *Transport Policy*, 50:106–114.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.
- Anable, J. and Gatersleben, B. (2005). All work and no play? The role of instrumental and affective factors in work and leisure journeys by different travel modes. *Transportation Research Part A: Policy and Practice*, 39:163–181.
- Axhausen, K. W., Bierlaire, M., and Abay, G. (2001). The acceptance of modal innovation. The case of Swissmetro. In Institut für Verkehrsplanung, Transporttechnik, S.-u. E. Z., editor, *1st Swiss Transport Research Conference*, volume 75, Zürich. Institut für Verkehrsplanung und Transporttechnik, Strassen- und Eisenbahnbau, ETH Zürich.
- Beirão, G. and Cabral, J. A. S. (2007). Understanding attitudes towards public transport and private car: A qualitative study. *Transport Policy*, 14:478–489.
- Ben-Akiva, M. and Boccara, B. (1995). Discrete choice models with latent choice sets. *International Journal of Research in Marketing*, 12(1):9–24.
- Ben-Akiva, M. and Lerman, S. (1985). Tests and Practical Issues in Developing Discrete Choice Models. In *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press series in transportation studies, chapter 7, pages 154–216. MIT Press, Cambridge.
- Ben-Akiva, M., McFadden, D., Train, K., Walker, J., Bhat, C., Bierlaire, M., Bolduc, D., Boersch-Supan, A., Brownstone, D., Bunch, D. S., Daly, A., Palma, A. D., Gopinath, D., Karlstrom, A., and Munizaga, M. A. (2002). Hybrid Choice Models: Progress and Challenges. *Marketing Letters*, 13(3):163–175.
- Berrett, T. B. and Samworth, R. J. (2021). USP: an independence test that improves on Pearson’s chi-squared and the G-test. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 477(2256):20210549.

- Bhat, C. R. (1995). A heteroscedastic extreme value model of intercity travel mode choice. *Transportation Research Part B: Methodological*, 29(6):471–483.
- Bierlaire, M. (2001). A general formulation of the cross-nested logit model. In *Swiss Transport Research Conference*, Monte Verita, Switzerland.
- Bierlaire, M. (2003). BIOGEME: a free package for the estimation of discrete choice models. In *3rd Swiss Transport Research Conference*.
- Bierlaire, M. (2023). A short introduction to Biogeme. Technical Report TRANSP-OR 230620, Transport and Mobility Laboratory, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland.
- Buehler, R. (2011). Determinants of transport mode choice: a comparison of Germany and the USA. *Journal of Transport Geography*, 19:644–657.
- Burman, P. (1989). A Comparative Study of Ordinary Cross-Validation, v-Fold Cross-Validation and the Repeated Learning-Testing Methods. *Biometrika*, 76(3):503–514.
- Burnham, K. P. and Anderson, D. R. (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, 33(2):261–304.
- Cantillo, V., Heydecker, B., and de Dios Ortúzar, J. (2006). A discrete choice model incorporating thresholds for perception in attribute values. *Transportation Research Part B: Methodological*, 40:807–825.
- Cervero, R. (2002). Built environments and mode choice: toward a normative framework. *Transportation Research Part D*, 7:265–284.
- Cosslett, S. R. (1981). Efficient Estimation of Discrete-Choice Models. In Manski, C. and McFadden, D., editors, *Structural Analysis of Discrete Data with Econometric Applications*, chapter 2, pages 51–111. The MIT Press.
- Dell’Olio, L., Ibeas, A., and Cecin, P. (2011). The quality of service desired by public transport users. *Transport Policy*, 18:217–227.
- Ding, C., Wang, D., Liu, C., Zhang, Y., and Yang, J. (2017). Exploring the influence of built environment on travel mode choice considering the mediating effects of car ownership and travel distance. *Transportation Research Part A: Policy and Practice*, 100:65–80.
- Golob, T. F. (2003). Structural equation modeling for travel behavior research. *Transportation Research Part B: Methodological*, 37(1):1–25.
- Greene, W. H. and Hensher, D. A. (2003). A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B: Methodological*, 37(8):681–698.
- Gumbel, E. J. (1958). *Statistics of Extremes*. Columbia University Press, New York.
- Hensher, D. A. and Greene, W. H. (2003). The Mixed Logit model: The state of practice. *Transportation*, 30(2):133–176.
- Hensher, D. A. and Ho, C. Q. (2016). Experience conditioning in commuter modal choice modelling – Does it make a difference? *Transportation Research Part E: Logistics and Transportation Review*, 95:164–176.

- Hess, S., Bierlaire, M., and Polak, J. W. (2005). Estimation of value of travel-time savings using mixed logit models. *Transportation Research Part A: Policy and Practice*, 39(2):221–236.
- Ho, C. Q., Hensher, D. A., Mulley, C., and Wong, Y. Z. (2018). Potential uptake and willingness-to-pay for Mobility as a Service (MaaS): A stated choice study. *Transportation Research Part A: Policy and Practice*, 117:302–318.
- Horeni, O., Gärling, T., Loukopoulos, P., and Fujii, S. (2007). An experimental simulation of adaptations to increased car-use costs. *Transportation Research Part F: Traffic Psychology and Behaviour*, 10:300–320.
- Johansson, M. V., Heldt, T., and Johansson, P. (2006). The effects of attitudes and personality traits on mode choice. *Transportation Research Part A: Policy and Practice*, 40:507–525.
- Kapitza, J. (2022). How people get to work at night. A discrete choice model approach towards the influence of nighttime on the choice of transport mode for commuting to work. *Journal of Transport Geography*, 104:103418.
- Long, J. (1997). Hypothesis Testing and Goodness of Fit. In *Regression Models for Categorical and Limited Dependent Variables*, Advanced Quantitative Techniques in the Social Sciences, chapter 4, pages 85–113. SAGE Publications.
- Mahani, S., Oliveira, B., Miguéis, V., Patrício, L., Carravilla, M., and Oliveira, J. (2025). Data-driven insights into transportation mode choices and their impact on sustainable mobility. Working paper.
- Masoumi, H. E. (2019). A discrete choice analysis of transport mode choice causality and perceived barriers of sustainable mobility in the MENA region. *Transport Policy*, 79:37–53.
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In Zarembka, P., editor, *Frontiers in Econometrics*, pages 105–142. Academic press, New York.
- McFadden, D. (1977). Modelling the Choice of Residential Location. *Transportation Research Record*.
- McFadden, D. and Train, K. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics*, 15(5):447–470.
- Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, 47(1):90–100.
- Nisbet, R., Miner, G., and Yale, K. (2018a). Data Understanding and Preparation. In Nisbet, R., Miner, G., and Yale, K., editors, *Handbook of Statistical Analysis and Data Mining Applications (Second Edition)*, chapter 4, pages 55–82. Academic Press, Boston, second edition.
- Nisbet, R., Miner, G., and Yale, K. (2018b). Model Evaluation and Enhancement. In Nisbet, R., Miner, G., and Yale, K., editors, *Handbook of Statistical Analysis and Data Mining Applications (Second Edition)*, chapter 11, pages 215–233. Academic Press, Boston, second edition.
- Parady, G., Ory, D., and Walker, J. (2021). The overreliance on statistical goodness-of-fit and under-reliance on model validation in discrete choice models: A review of validation practices in the transportation academic literature. *Journal of Choice Modelling*, 38:100257.

- Ratner, B. (2009). The correlation coefficient: Its values range between $+1/-1$, or do they? *Journal of Targeting, Measurement and Analysis for Marketing*, 17.
- Rodríguez, D. A. and Joo, J. (2004). The relationship between non-motorized mode choice and the local physical environment. *Transportation Research Part D: Transport and Environment*, 9:151–173.
- Romão, J. and Bi, Y. (2021). Determinants of collective transport mode choice and its impacts on trip satisfaction in urban tourism. *Journal of Transport Geography*, 94:103094.
- Rubin, D. B. (1976). Inference and Missing Data. *Biometrika*, 63(3):581–592.
- Salon, D. (2009). Neighborhoods, cars, and commuting in New York City: A discrete choice approach. *Transportation Research Part A: Policy and Practice*, 43:180–196.
- Saneinejad, S., Roorda, M. J., and Kennedy, C. (2012). Modelling the impact of weather conditions on active transportation travel behaviour. *Transportation Research Part D: Transport and Environment*, 17(2):129–137.
- Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461–464.
- Sifringer, B., Lurkin, V., and Alahi, A. (2020). Enhancing discrete choice models with representation learning. *Transportation Research Part B: Methodological*, 140:236–261.
- St-Louis, E., Manaugh, K., Lierop, D. V., and El-Geneidy, A. (2014). The happy commuter: A comparison of commuter satisfaction across modes. *Transportation Research Part F: Traffic Psychology and Behaviour*, 26:160–170.
- Ton, D., Duives, D. C., Cats, O., Hoogendoorn-Lanser, S., and Hoogendoorn, S. P. (2019). Cycling or walking? Determinants of mode choice in the Netherlands. *Transportation Research Part A: Policy and Practice*, 123:7–23.
- Train, K. E. (2009a). Bayesian procedures. In *Discrete Choice Methods with Simulation*, chapter 12, pages 321–356. Cambridge University Press.
- Train, K. E. (2009b). Gev. In *Discrete Choice Methods with Simulation*, chapter 4, pages 87–110. Cambridge University Press.
- Train, K. E. (2009c). Introduction. In *Discrete Choice Methods with Simulation*, chapter 1, pages 1–12. Cambridge University Press.
- Train, K. E. (2009d). Logit. In *Discrete Choice Methods with Simulation*, chapter 3, pages 34–75. Cambridge University Press.
- Train, K. E. (2009e). Mixed logit. In *Discrete Choice Methods with Simulation*, chapter 6, pages 153–172. Cambridge University Press.
- Train, K. E. (2009f). Properties of discrete choice models. In *Discrete Choice Methods with Simulation*, chapter 2, pages 15–40. Cambridge University Press.
- Train, K. E. (2009g). Variations on a theme. In *Discrete Choice Methods with Simulation*, chapter 7, pages 173–208. Cambridge University Press.
- Tyrinopoulos, Y. and Antoniou, C. (2013). Factors affecting modal choice in urban mobility. *European Transport Research Review*, 5:27–39.

- van Buuren, S. and Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, 45(3):1–67.
- Vovsha, P. (1997). Application of Cross-Nested Logit Model to Mode Choice in Tel Aviv, Israel, Metropolitan Area. *Transportation Research Record*, 1607(1):6–15.
- Vuong, Q. H. (1989). Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica*, 57(2):307–333.
- Walker, J. and Ben-Akiva, M. (2002). Generalized random utility model. *Mathematical Social Sciences*, 43(3):303–343.
- Wardman, M., Tight, M., and Page, M. (2007). Factors influencing the propensity to cycle to work. *Transportation Research Part A: Policy and Practice*, 41:339–350.
- Wen, C.-H. and Koppelman, F. S. (2001). The generalized nested logit model. *Transportation Research Part B: Methodological*, 35(7):627–641.
- Whalen, K. E., Páez, A., and Carrasco, J. A. (2013). Mode choice of university students commuting to school and the role of active travel. *Journal of Transport Geography*, 31:132–142.
- Williams, H. C. W. L. (1977). On the Formation of Travel Demand Models and Economic Evaluation Measures of User Benefit. *Environment and Planning A: Economy and Space*, 9(3):285–344.
- Wu, H. and Leung, S.-O. (2017). Can Likert Scales be Treated as Interval Scales?—A Simulation Study. *Journal of Social Service Research*, 43(4):527–532.
- Zhou, J. (2012). Sustainable commute in a car-dominant city: Factors affecting alternative mode choices among university students. *Transportation Research Part A: Policy and Practice*, 46:1013–1029.

Appendix A

Descriptive statistics

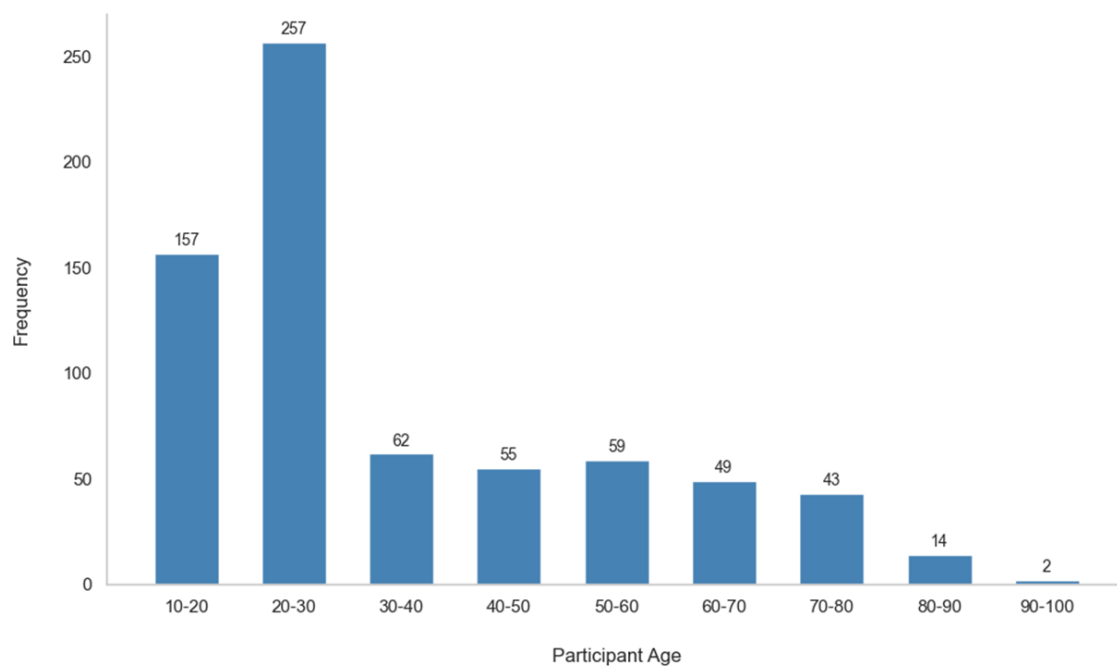


Figure A.1: Univariate distribution of participant's age.

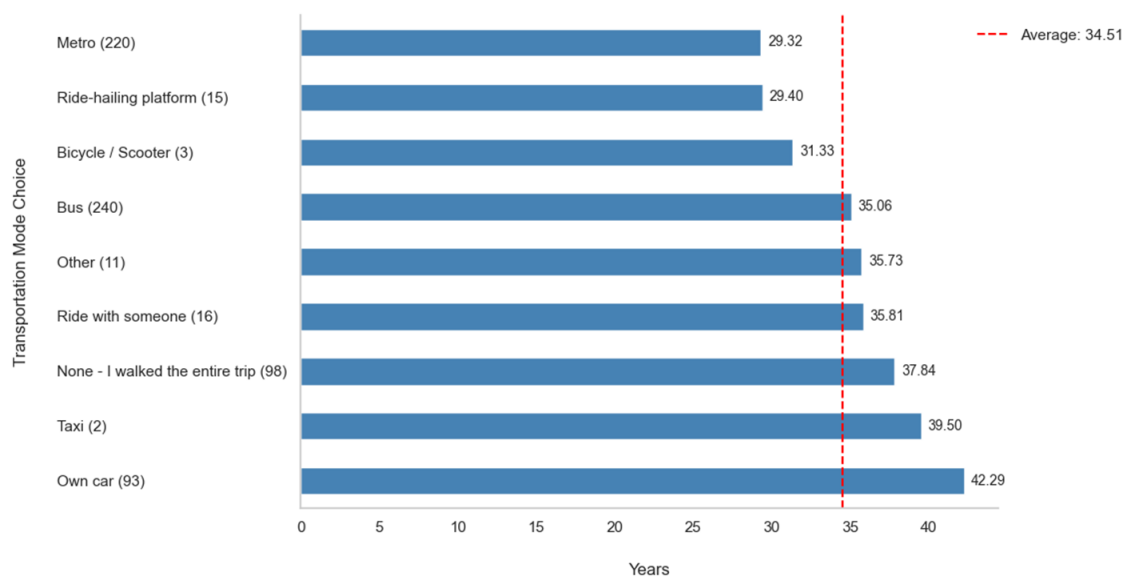


Figure A.2: Distribution of average participant's age, by mode choice.

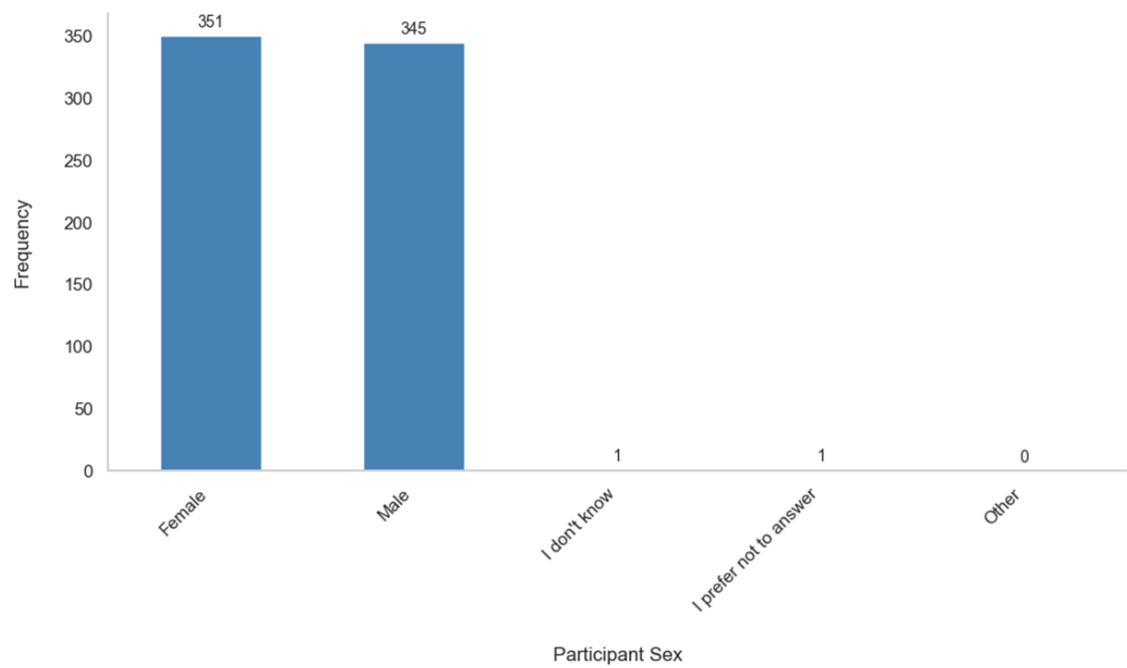


Figure A.3: Univariate distribution of participant's sex.

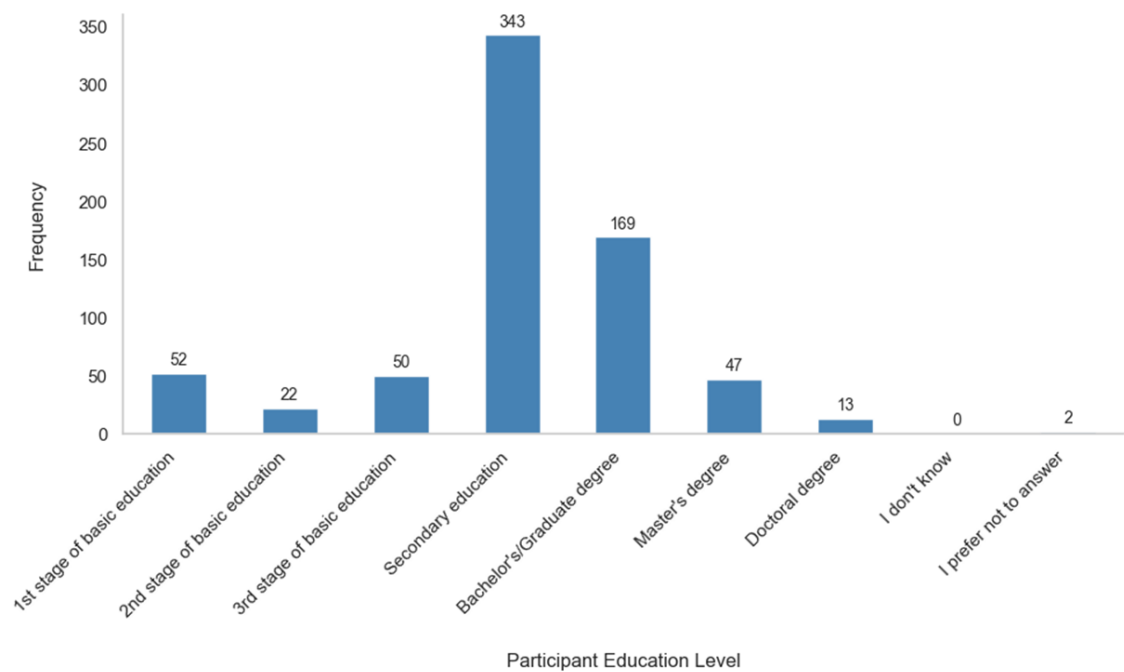


Figure A.4: Univariate distribution of participant's education level.

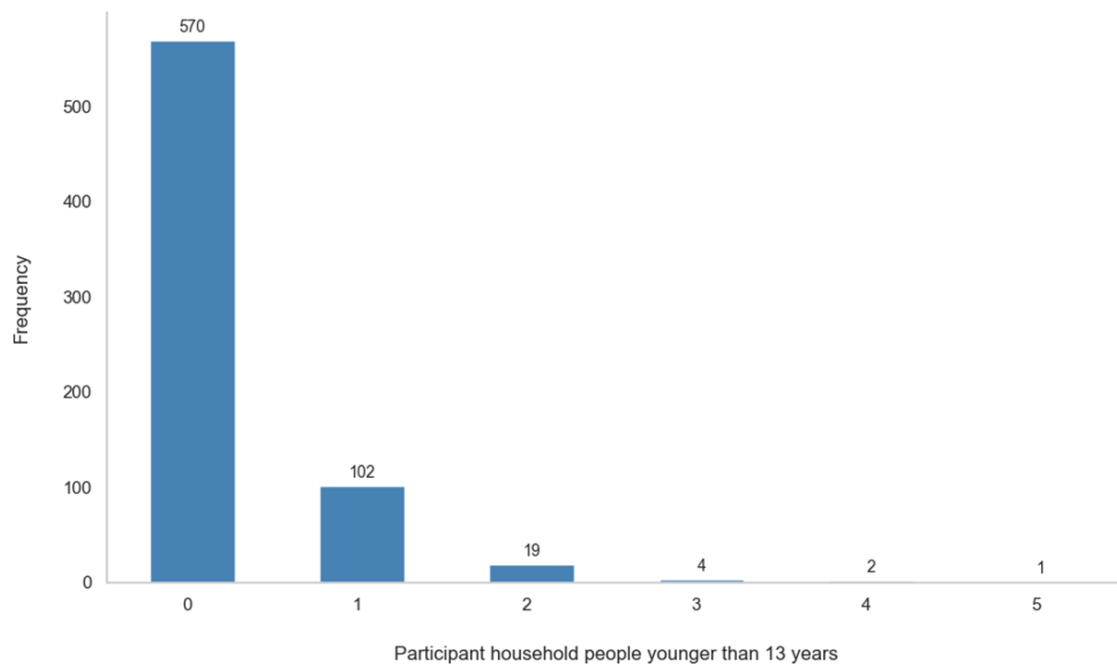


Figure A.5: Univariate distribution of participant's household composition.

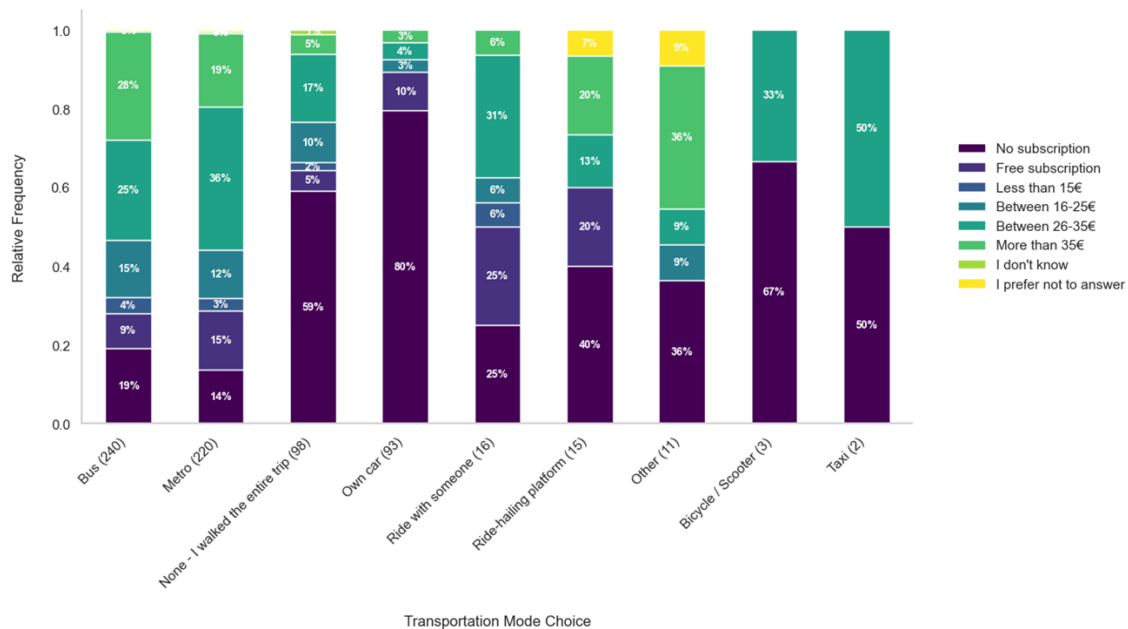


Figure A.6: Distribution of participant's subscription owned, by mode choice.

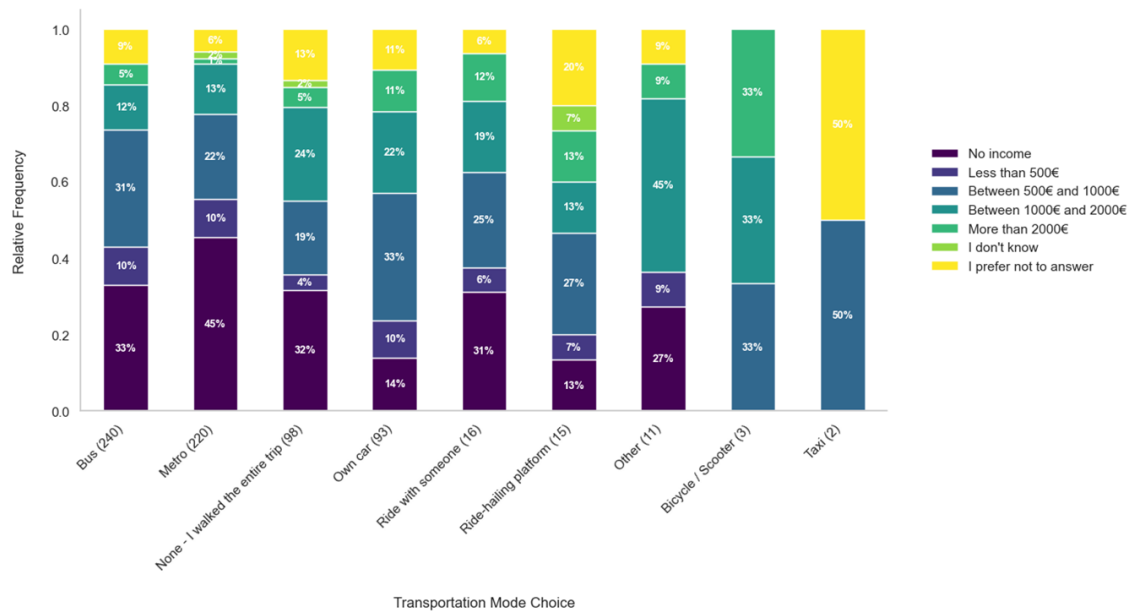


Figure A.7: Distribution of participant's monthly income level, by mode choice.

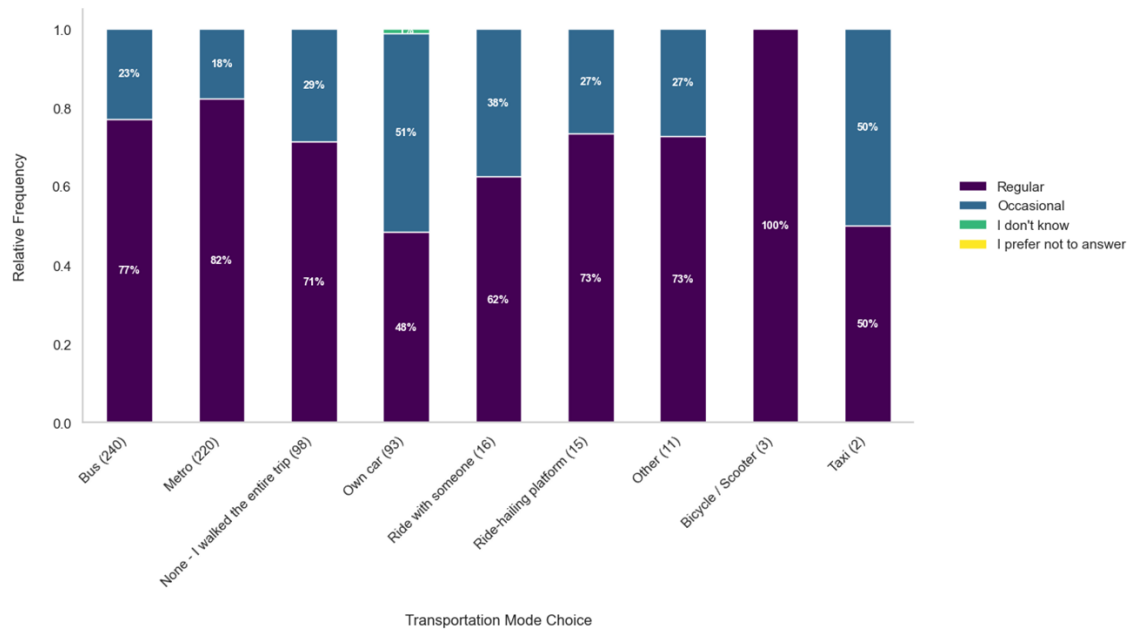


Figure A.8: Distribution of whether the participant's trip was regular, by mode choice.

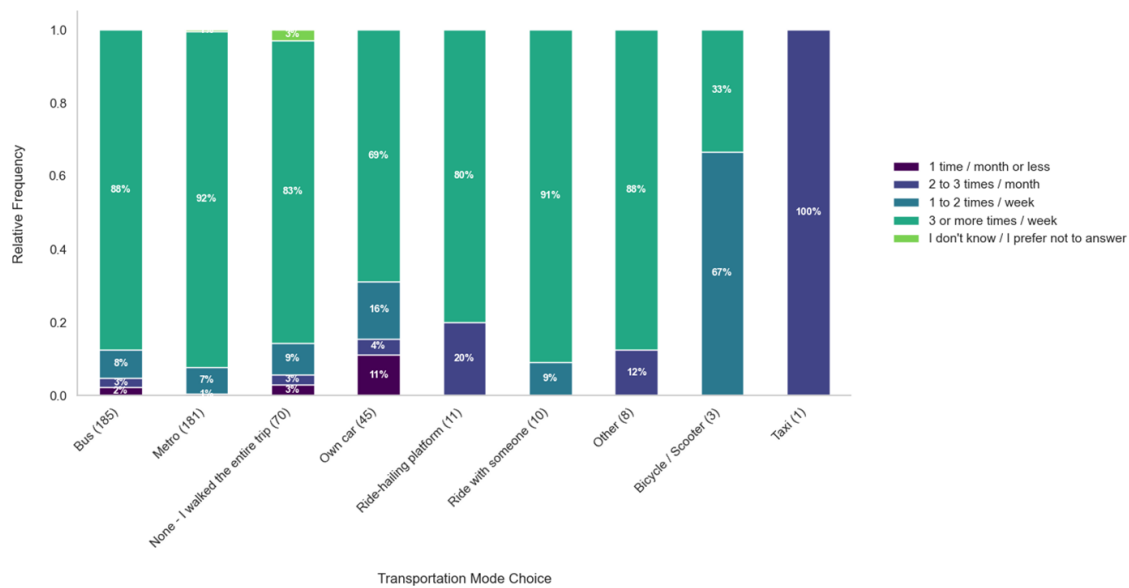


Figure A.9: Distribution of participant's trip regularity, by mode choice.

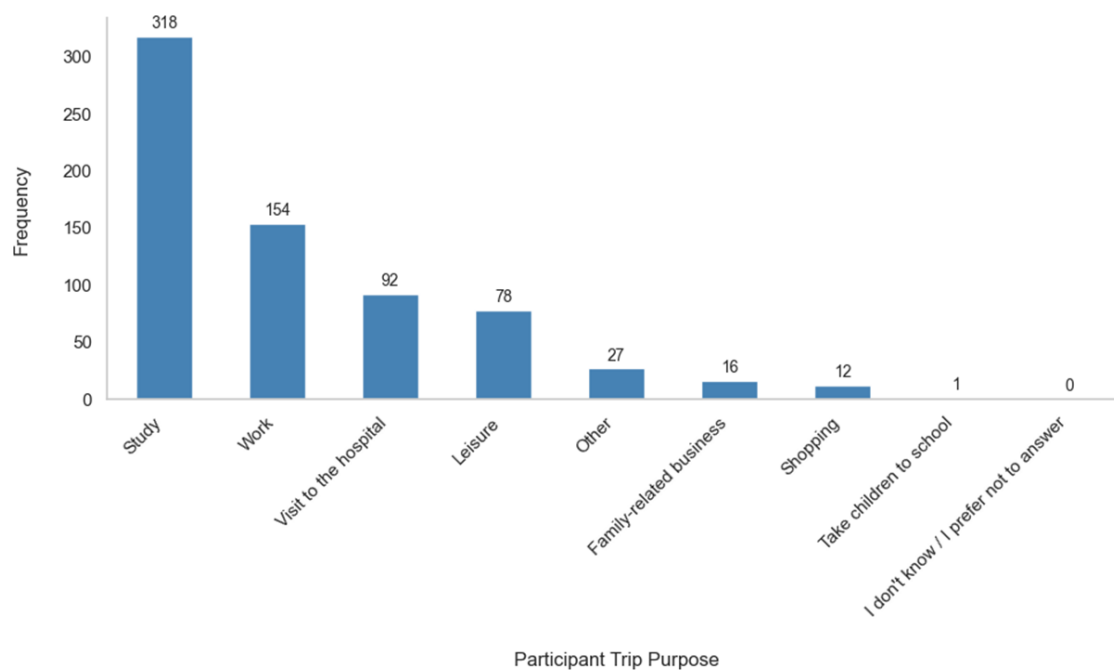


Figure A.10: Univariate distribution of participant's trip purpose.

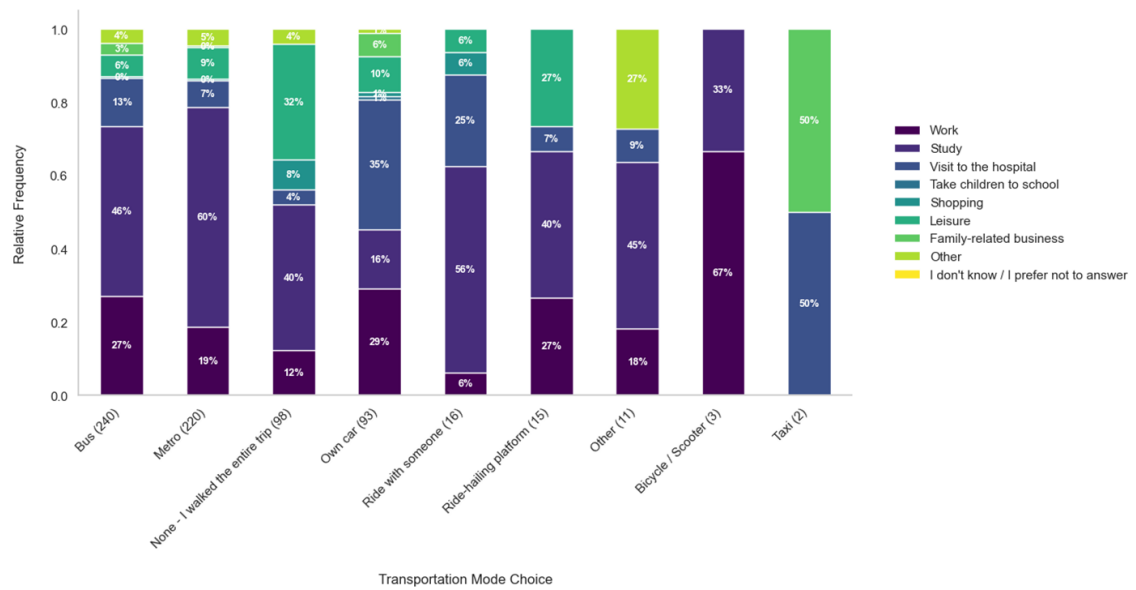


Figure A.11: Distribution of participant's trip purpose, by mode choice.

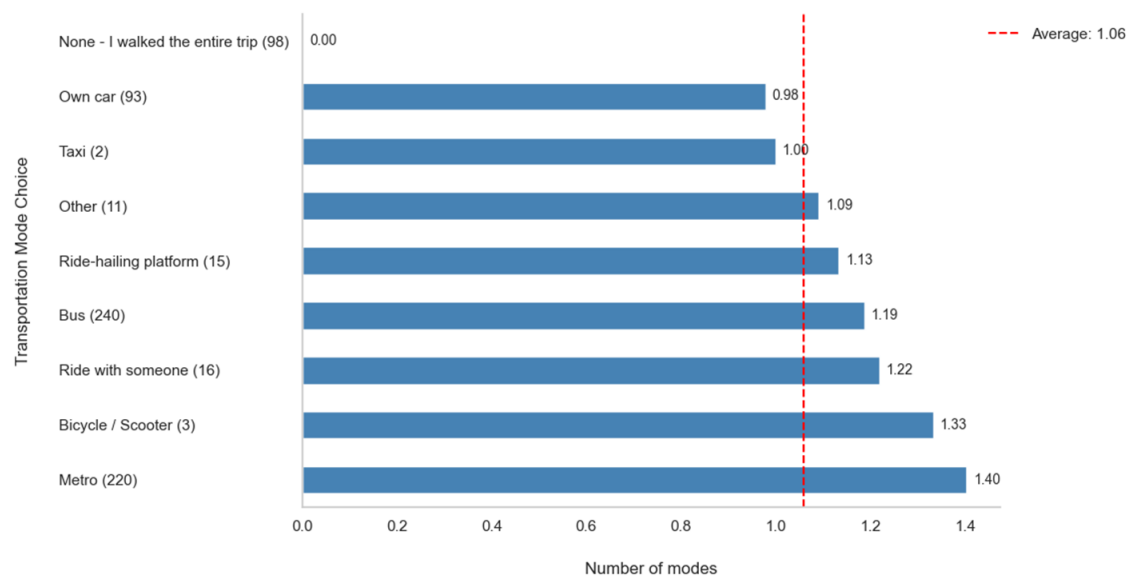


Figure A.12: Distribution of average number of modes used, by mode choice.

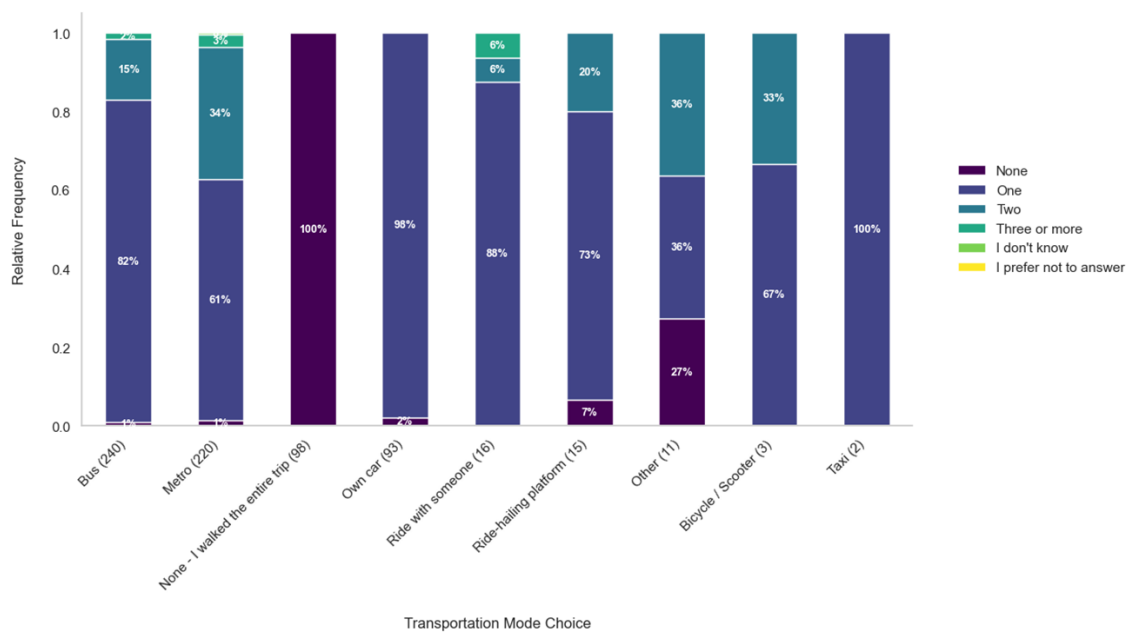


Figure A.13: Distribution of number of modes used, by mode choice.

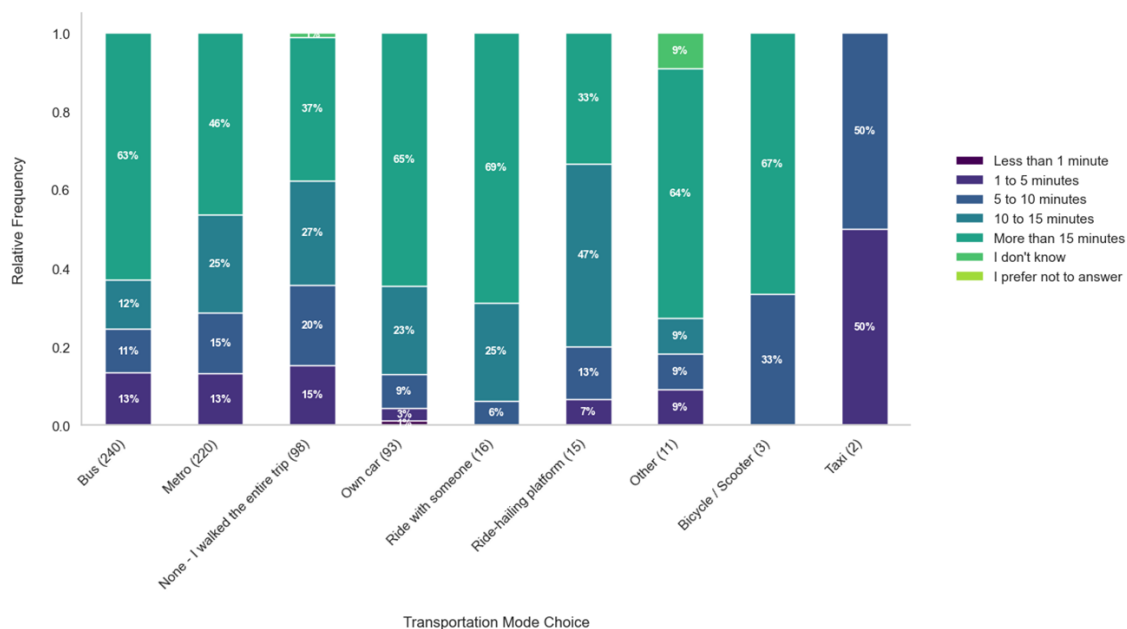


Figure A.14: Distribution of trip duration, by mode choice.

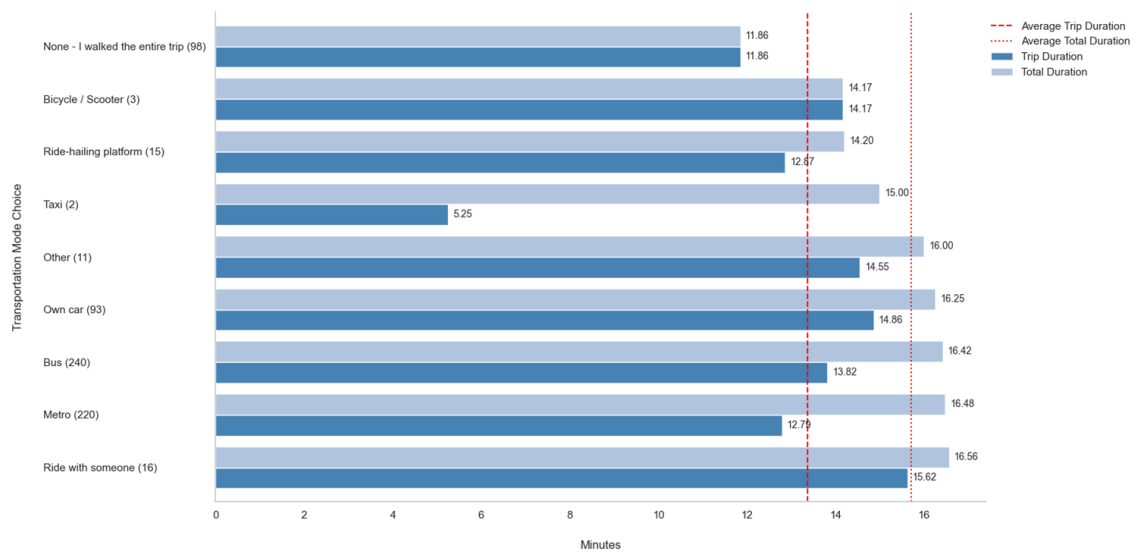


Figure A.15: Distribution of average trip duration vs. total journey duration, by mode choice.

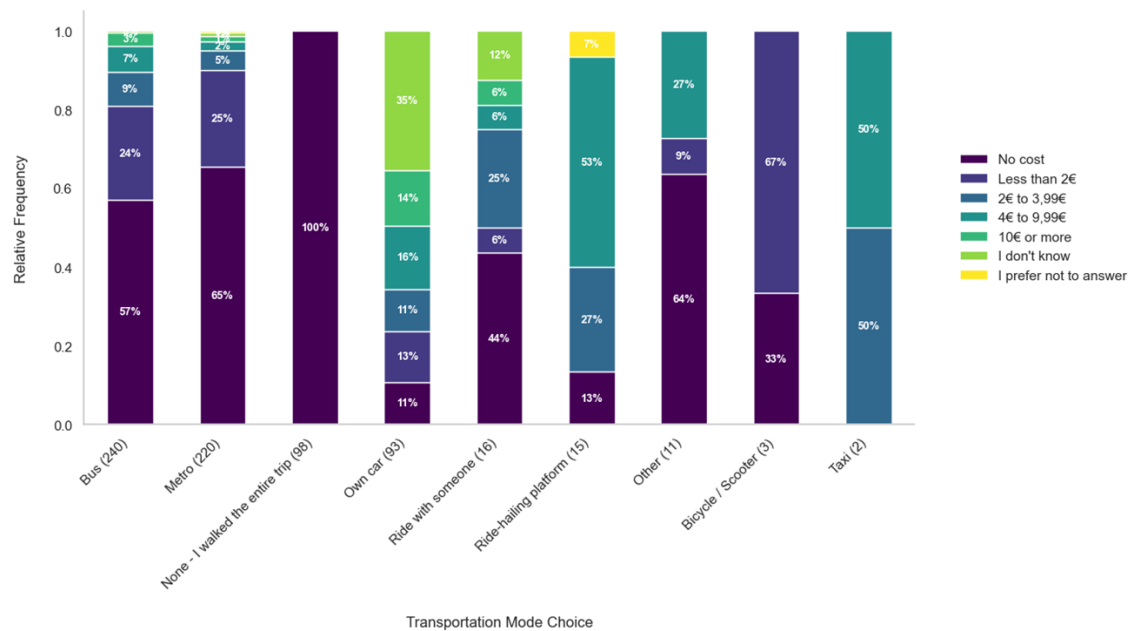


Figure A.16: Distribution of trip cost, by mode choice.

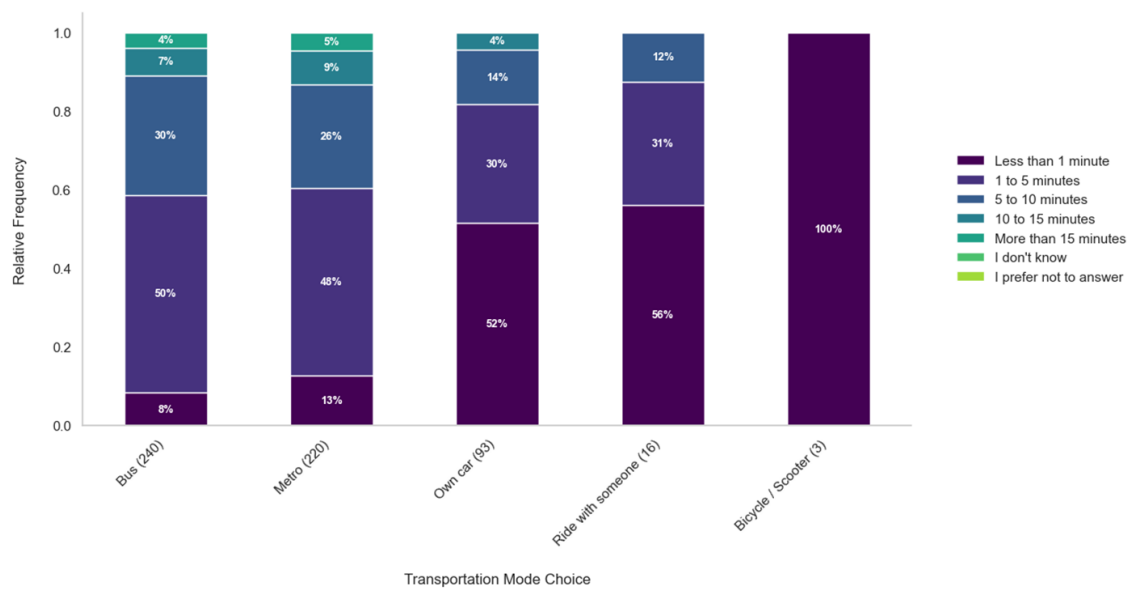


Figure A.17: Distribution of walking time, by mode choice.

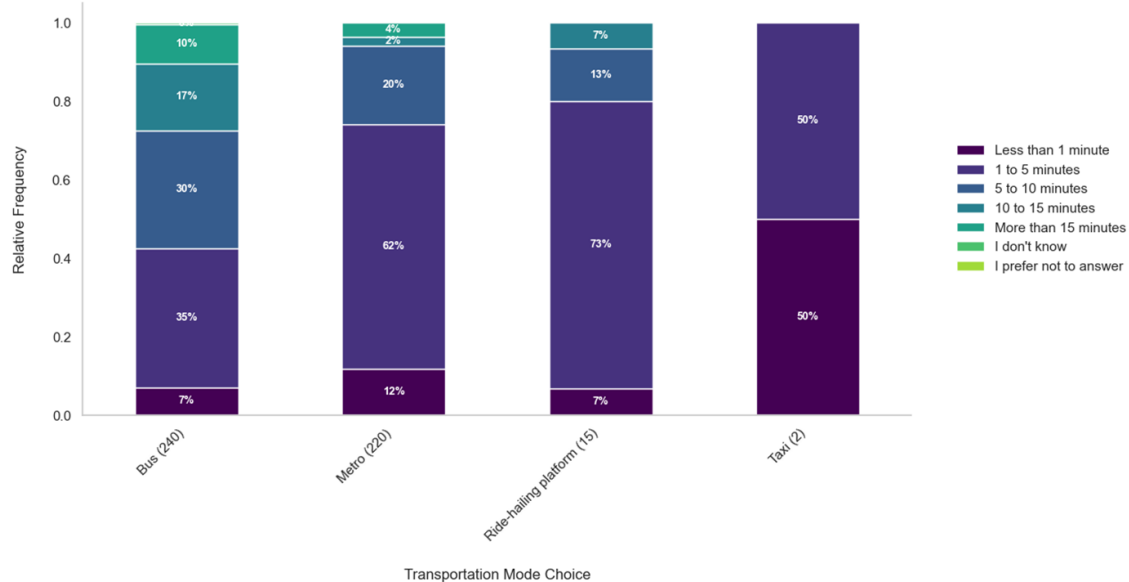


Figure A.18: Distribution of waiting time, by mode choice.

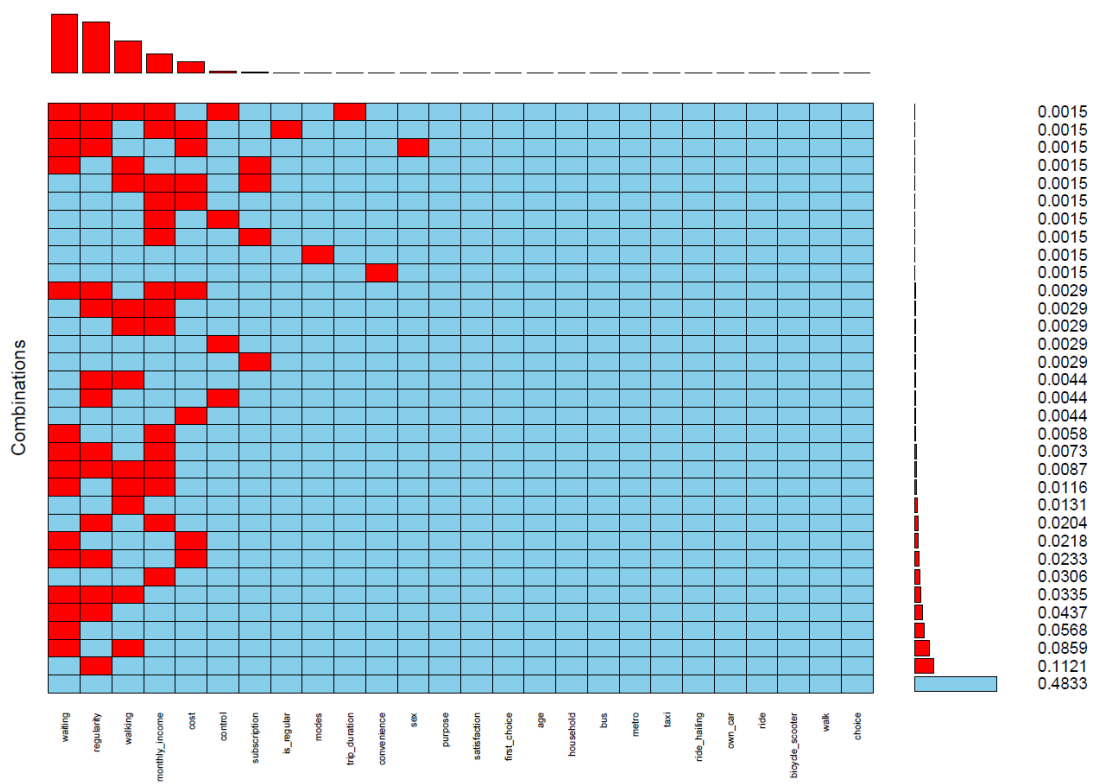


Figure A.19: Pattern of missing data.

Appendix B

Parameter estimation results

Table B.1: Parameter estimation results from LCM - Class 1 using set of factors "Threshold test step 2". The estimated value of H_{n1} is 0.397. Coefficients are scaled by a factor of 0.1 for compact representation.

Factor	Bus ($N = 240$)	Metro ($N = 220$)	Walk ($N = 98$)	Own Car ($N = 93$)	Riding with someone ($N = 16$)	Ride-hailing ($N = 15$)	Bicycle/Scooter ($N = 3$)	Taxi ($N = 2$)
ASC	0.00 ^a —	20.00*** (5.85)	0.86*** (7.86)	0.14 (1.58)	0.69*** (5.88)	-8.77*** (-9.61)	0.22*** (6.81)	-0.02*** (-13.80)
<i>Socio-demographic</i>								
Owens a subscription: yes	33.10*** (10.40)	-9.69*** (-10.90)	-10.10*** (-10.80)	-7.92*** (-10.00)	3.51*** (12.00)	-2.17*** (-5.86)	-6.75*** (-9.72)	-0.002 (0.00)
Owens a subscription: no ^b	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Income level: high	-14.50*** (-10.40)	-5.85*** (-11.30)	1.81*** (9.28)	4.06*** (12.70)	-0.82*** (-8.27)	15.10*** (10.20)	0.20*** (2.91)	-0.00 (-0.00)
Income level: low	-42.60*** (-9.99)	16.20*** (9.31)	4.39*** (9.73)	0.08 (0.90)	-0.48*** (-9.17)	21.40*** (10.50)	1.06*** (10.80)	-0.02*** (-8.75)
Income level: none ^b	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>Objective</i>								
Trip regularity	-4.30*** (-9.62)	1.72*** (12.10)	0.40*** (4.52)	3.03*** (10.20)	2.55*** (11.10)	-3.50*** (-11.00)	0.15** (2.20)	-0.04*** (-16.00)
Number of modes	23.30*** (10.40)	-13.70*** (-9.27)	^c	2.16*** (8.63)	0.30*** (3.70)	44.40*** (10.30)	8.34*** (11.10)	-0.02*** (-10.90)
Trip duration	-3.96*** (-10.10)	0.13*** (3.30)	1.58*** (9.32)	0.64*** (8.25)	3.91*** (11.10)	-3.58*** (-9.84)	1.65*** (10.40)	-0.37*** (-12.10)
Cost	-7.04*** (-10.30)	-5.65*** (-11.00)	^c	11.00*** (11.20)	1.58*** (12.20)	6.71*** (11.00)	0.20*** (6.02)	-0.00** (-2.40)
Walking time	18.30*** (9.96)	22.50*** (10.10)	^c	-16.10*** (-10.50)	0.55*** (3.18)	^c	-18.50*** (-10.50)	^c
Waiting time	25.20*** (10.10)	11.80*** (10.10)	^c	^c	^c	13.10*** (10.70)	^c	-0.05*** (-10.40)
<i>Subjective</i>								
Satisfaction	-10.40*** (-9.94)	-3.28*** (-5.23)	3.12*** (9.86)	0.62** (2.33)	0.60*** (2.64)	6.49*** (9.16)	2.96*** (8.48)	-0.11*** (-15.00)
Perception as first choice	3.80*** (9.61)	-4.73*** (-9.67)	1.41*** (13.60)	4.80*** (9.12)	-6.11*** (-10.70)	-1.91*** (-9.52)	2.81*** (10.00)	-0.07*** (-18.60)

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$

t-statistics in parentheses.

"Ride" stands for riding with someone.

^a Bus is the reference alternative, therefore its constant is normalized to zero.

^b Reference category for the respective factor, therefore all coefficients are normalized to zero.

^c Not applicable for the respective alternative.

Table B.2: Parameter estimation results from LCM - Class 2 using set of factors "Threshold test step 2". By inference, the value of H_{n2} is 0.603. Coefficients are scaled by a factor of 0.1 for compact representation.

Factor	Bus ($N = 240$)	Metro ($N = 220$)	Walk ($N = 98$)	Own Car ($N = 93$)	Riding with someone ($N = 16$)	Ride-hailing ($N = 15$)	Bicycle/Scooter ($N = 3$)	Taxi ($N = 2$)
ASC	0.00 ^a —	41.80*** (14.10)	10.30*** (19.60)	78.80*** (15.80)	12.10*** (21.10)	-6.09*** (-20.70)	-0.65*** (-14.50)	1.55 (0.00)
<i>Socio-demographic</i>								
Owens a subscription: yes	18.60*** (17.20)	51.40*** (19.30)	-62.80*** (-19.40)	-15.90*** (-16.50)	5.78*** (13.00)	2.73*** (105.00)	-0.62*** (-13.80)	0.75 (0.00)
Owens a subscription: no ^b	0.00 —	0.00 —	0.00 —	0.00 —	0.00 —	0.00 —	0.00 —	0.00 —
Income level: high	-33.50*** (-20.20)	-14.00*** (-22.70)	17.00*** (19.50)	-0.42 (-0.39)	30.80*** (22.40)	-1.37 (-0.00)	-0.06 (-0.00)	1.58 (0.00)
Income level: low	-1.69 (-0.00)	-23.90*** (-19.20)	33.20*** (19.00)	-8.18*** (-18.00)	-15.10*** (-20.20)	10.90*** (22.20)	3.88*** (19.50)	0.97 (0.00)
Income level: none ^b	0.00 —	0.00 —	0.00 —	0.00 —	0.00 —	0.00 —	0.00 —	0.00 —
<i>Objective</i>								
Trip regularity	-3.80*** (-15.10)	0.05 (0.24)	-0.93*** (-4.69)	8.73*** (17.20)	-5.39*** (-12.30)	14.50*** (20.80)	-12.30*** (-19.50)	-0.88 (-0.00)
Number of modes	-0.39 (-1.36)	72.90*** (18.80)	^c —	-9.53*** (-17.30)	6.00*** (13.80)	9.20*** (23.20)	3.93*** (19.80)	2.41*** (20.10)
Trip duration	2.95*** (19.10)	-7.31*** (-17.80)	2.09*** (19.60)	-2.49*** (-19.60)	-1.09*** (-11.20)	0.70*** (13.50)	5.23*** (18.80)	-0.07*** (-3.63)
Cost	12.80*** (19.80)	-8.16*** (-13.90)	^c —	13.80*** (19.60)	12.40*** (18.80)	24.30*** (20.10)	3.73*** (18.80)	13.90*** (19.40)
Walking time	17.60*** (20.30)	19.70*** (20.20)	^c —	13.30*** (20.50)	6.49*** (17.00)	^c —	-0.63*** (-28.30)	^c —
Waiting time	9.96*** (18.30)	12.20*** (18.40)	^c —	^c —	^c —	8.14*** (23.50)	^c —	2.80*** (19.30)
<i>Subjective</i>								
Satisfaction	-11.50*** (-19.40)	-29.50*** (-19.00)	11.30*** (18.20)	3.29*** (11.70)	19.10*** (18.50)	-3.05*** (-27.70)	3.46*** (17.90)	6.84*** (22.10)
Perception as first choice	2.70*** (21.20)	14.00*** (19.60)	6.91*** (19.60)	0.04 (0.00)	2.50*** (18.90)	-30.00*** (-20.20)	1.31*** (8.58)	2.57 (0.00)

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$

t -statistics in parentheses.

^a Bus is the reference alternative, therefore its constant is normalized to zero.

^b Reference category for the respective factor, therefore all coefficients are normalized to zero.

^c Not applicable for the respective alternative.