Politecnico di Torino

Master of Science in Civil Engineering

A.Y. 2025/2026

Modelling Mode Choice Behaviour by Trip Purpose in the Context of the 15-Minute City: an Application to a Suburban Area in Lisbon (Portugal)

Supervisors:

Marco Diana

João de Abreu e Silva

Candidate:

Samuele Milani

July 2026 Term

Ai miei genitori e a mia sorella, che hanno continuato a credere in me dopo i primi due "scricchiolanti" anni di percorso.

Ai miei nonni, che mi hanno sempre fatto sentire orgoglioso di essere la persona che sono.

Agli amici di Torino, Pisa e Lisbona, così presenti nella mia vita da diventare una seconda famiglia.

ABSTRACT

The 15-minute city has recently gained attention as a planning approach by promoting active mobility through proximity to essential amenities. Existing literature has predominantly focused on dense urban cores and commuting patterns, frequently neglecting peripheral contexts and non-mandatory travel. This thesis addresses these subjects by investigating the determinants of mode choice in "Vila Franca de Xira", a suburban municipality in the Lisbon Metropolitan Area.

Based on an in-person intercept survey of more than a thousand respondents conducted in May 2025, the research adopts an advanced discrete choice modelling approach to analyse travel behaviour. The study compares a traditional Multinomial Logit (MNL) model with a more complex Hybrid Choice Model (or ICLV - Integrated Choice and Latent Variable Model). The latter explicitly integrates two latent variables identified as "Attitude towards the 15-minute City" and "Perceived Easiness of Car Use" to capture the unobserved psychological constructs influencing user decisions.

The research design differentiates the analysis of mandatory trips (commuting) from non-mandatory ones (leisure, visiting, shopping), a distinction that was overlooked in previous research in this area. Results highlight that, while commuting patterns appear more rigid and car-oriented, discretionary trips are significantly influenced by psychosocial attitudes and quality of the urban environment. Furthermore, various differences are highlighted in the models across socio-demographic factors, particularly regarding the role of age and education level.

These findings, coupled with the improved behavioural interpretation given by the latent variables, confirm the effectiveness of the ICLV models in understanding mobility patterns especially for discretionary trips in the context of the 15-minute city. Additionally, they suggest that policies should rely not only on physical infrastructure but also on the psychological perceptions of users, particularly for non-mandatory mobility.

RIASSUNTO

Scopo dello spostamento e scelta modale nel contesto della città dei 15 minuti: il caso studio di un'area suburbana di Lisbona (Portogallo)

La città dei 15 minuti ha acquisito rilevanza in tempi recenti come approccio urbanistico volto a promuovere la mobilità attiva e la prossimità ai servizi essenziali. Sinora, la letteratura esistente si è concentrata prevalentemente su contesti urbani ad alta densità e su spostamenti pendolari, trascurando spesso le aree periferiche e i viaggi non sistematici. Questa tesi affronta tali aspetti analizzando i determinanti della scelta modale nel comune suburbano di "Vila Franca de Xira", situato nell'Area Metropolitana di Lisbona.

Facendo riferimento a un'indagine sul campo condotta nel maggio 2025 su oltre mille intervistati, il presente studio adotta un approccio modellistico avanzato di scelta discreta in modo tale da analizzare le abitudini di spostamento degli abitanti. Lo studio confronta un tradizionale modello Logit Multinomiale (MNL) con un più complesso Modello Ibrido a Variabili Latenti (ICLV). Quest'ultimo integra due variabili latenti identificate come "Attitudine nei confronti della città dei 15 minuti" e "Agio del trasporto privato", al fine di cogliere i costrutti psicologici non osservabili che influenzano le decisioni degli utenti.

Nella presente ricerca vengono distinti gli spostamenti sistematici (pendolarismo) da quelli non sistematici (tempo libero, visite e shopping). I risultati evidenziano come i comportamenti legati al pendolarismo risultino più rigidi e orientati all'uso dell'automobile, mentre gli spostamenti discrezionali siano maggiormente influenzati dagli atteggiamenti psicosociali e dalla qualità dell'ambiente urbano. Inoltre, emergono differenze tra i diversi modelli riguardo a fattori sociodemografici, in particolare per quanto concerne il ruolo dell'età e del livello di istruzione.

Questi risultati, uniti alla maggiore interpretabilità comportamentale dovuta alle variabili latenti, confermano l'utilità dei modelli ICLV per questo approccio analitico. Inoltre, essi suggeriscono che le decisioni prese dai politici per promuovere la mobilità sostenibile non dovrebbero basarsi esclusivamente sulle infrastrutture fisiche, ma anche sulle percezioni e sugli atteggiamenti psicologici degli utenti, in particolare per quanto riguarda la mobilità non sistematica.

STATEMENT ON THE USE OF GENERATIVE ARTIFICIAL INTELLIGENCE TOOLS

During the preparation of this thesis, Google Gemini 3.1 Pro was employed as a support instrument during selected phases of the research and writing process. Here is a detailed list of tasks performed with AI assistance:

- Editorial support: The tool was principally used for correcting grammatical errors and paraphrasing sentences to avoid repetitions.
- Literature review: AI was used for searching papers and extracting relevant information from sources. The latter were then independently verified and critically reviewed by the author.

The output and refinements provided by this tool are primarily contained in the following sections of the work:

- Abstract
- Chapter 1 (Introduction)
- Chapter 2 (Literature Review)
- Chapter 5 (Conclusions)

All AI-generated content was critically reviewed, fact-checked, and edited to align with academic rigor. Citations and arguments were independently verified by the candidate. Therefore, the candidate takes full responsibility for the accuracy, originality and integrity of the final work.

Table of contents

List of Tables

List of Figures

1. Introduction

In recent years, increasing attention has been addressed to urban planning models capable of promoting sustainable mobility and reducing car dependency. The concept of 15-minute city has emerged rapidly in this context, aiming to provide access to services and activities by active modes within a short range. By encouraging unmotorized modes and reducing the need for long car trips, this framework has established itself as one of the most prominent in transport planning.

Most studies on the 15-minute city have primarily focused on dense urban environments (Arias-Molinares et al., 2025), where accessibility conditions and amenities are already developed. In contrast, suburban and peripheral areas often present different spatial and mobility characteristics, typically characterised by lower density, greater dependence on private vehicles and weaker accessibility to essential services. For this reason, the applicability of proximity-based urban models in suburban contexts is often considered a conflicting topic.

At the same time, existing research is largely concentrated on mandatory trips and has often neglected discretionary ones (Gim, 2017). Although commuting represents a significant component of daily travel demand, non-mandatory trips such as shopping, leisure and visiting activities also play an important role in shaping urban mobility behaviour. These trips are generally more flexible and potentially more sensitive to hidden behavioural and perception factors. Nevertheless, the interaction between trip purposes, mode choice and behavioural attitudes remains relatively unexplored in the literature, especially in suburban contexts.

Another relevant limitation of current research concerns the treatment of psychological and attitudinal dimensions influencing transport behaviour (Vij et al., 2013). The majority of current studies relies on traditional discrete choice models, such as MNL (Multinomial Logit). These are generally based on observable variables like socio-demographic characteristics and level of service attributes, while latent behavioural factors are often ignored. However, perceptions, habits and attitudes towards urban mobility could significantly influence transport choices, particularly when analysing discretionary travel behaviour. Through specific models, such as ICLV (Integrated Choice and Latent Variable), it is possible to capture these hidden constructs, increasing the overall explanatory power of the framework.

This thesis addresses these gaps by investigating transport mode choice in the suburban municipality of "Vila Franca de Xira", located in the LMA (Lisbon Metropolitan Area). Particular attention is awarded to the distinction between mandatory and non-mandatory trips, with the objective of understanding if different trip purposes are effectively associated with different behavioural mechanisms.

The study adopts a framework based on the comparison between a traditional MNL model and an ICLV model, a subtype of HCM (Hybrid Choice Models). In addition to observable socio-demographic variables, the specification of the latter will incorporate two latent constructs. One will be related to attitudes towards 15-minute city and the other will evaluate the perception of private transport comfort. The objective is to assess whether the greater explanatory power of the second approach justifies the increase in model complexity and data gathering effort.

Overall, this thesis contributes to the existing literature by integrating trip purpose analysis, transport mode choice modelling and latent behavioural dimensions within a suburban context. In doing so, it provides an analysis of the applicability of the 15-minute city model beyond dense urban centres and highlights the importance of considering both infrastructural and psychological factors in sustainable mobility policies.

The workflow is divided into five interconnected chapters. Following this introduction, the thesis proceeds with a literature review on the 15-minute city, transport mode choice and modelling approaches. Chapter 3 describes the methodology, focusing on the study area, the survey design and the statistical methods employed to specify the models. The subsequent part presents the modelling results, while chapter 5 discusses the main outcomes, limitations and directions for future research.

2. Literature Review

2.1 The Paradigm of the 15-Minute City

The concept of the "15-minute city" (*la ville du quart d'heure* in French) represents a paradigm shift in contemporary urban planning, challenging the traditional zoning models that dominated the 20th century. The historic focus on motorized mobility is being progressively abandoned in favour of the promotion of active modes, both for environmental and social purposes.

First conceptualized by Carlos Moreno (2016), the model proposes a new type of urban planning, where the six essential social functions (living, working, supplying, caring, learning and enjoying) are accessible within a 15-minutes time limit via soft mobility, specifically walking or cycling.

2.1.1 From Mobility to Accessibility: Chrono-Urbanism

At the core of this approach lies the concept of "chrono-urbanism," which suggests that urban quality of life is inversely proportional to the time spent on forced mobility. Unlike traditional transport models oriented towards speed and flow maximization, the 15-minute city aims for "de-motorization" through a spatial reorganization of amenities. Although its practical application gained global resonance with Paris Mayor Anne Hidalgo in 2020 as a response to COVID-19 pandemic, its theoretical foundation is rooted in the principles of "Proximity, Diversity, Density, and Digitalization" (Moreno et al., 2021). The objective is not only to reduce emissions, but to restore the value of time to citizens, transforming travel from a logistical necessity into an enjoyable urban experience.

2.1.2 The Gap between Proximity and Mode Choice

Scientific literature warns against a too simplistic relation between distance and unmotorized preference. While "Distance Decay" functions in accessibility modelling (Martínez & Viegas, 2013) suggest that reducing spaces increases the marginal utility of active modes, other studies demonstrate that spatial proximity is a necessary but insufficient condition for behavioural change (Ewing & Cervero, 2010). Further research highlights that, even in dense urban contexts where services are accessible within 15 minutes, a significant percentage of residents continue to use private cars due to cultural habits (Murooka et al., 2025) or perception of insecurity (Loukaitou-Sideris, 2006). This phenomenon suggests the existence of other barriers that models based purely on density fail to capture. For this reason, it is important to investigate the latent psychological constructs that are triggering the real shift to active mobility.

2.1.3 Psychosocial Determinants and Infrastructure

On the basis of the previous discussion, we can conclude that the success of the 15-minute city depends on satisfying complex hierarchical needs. Referring to the "Hierarchy of Walking Needs" (Alfonzo, 2005), the decision to walk is not triggered only by feasibility or proximity, but requires higher levels of accessibility, safety, comfort, and pleasurability. Furthermore, psychosocial factors such as perceived social norms and individual attitudes towards the urban environment are starting to play a crucial role. Recent literature confirms that, while physical infrastructure is a prerequisite, it is not the only determinant of active mobility. A behavioural study conducted on young adults (Verhoeven et al., 2016) demonstrate that psychosocial factors, specifically social support and perceived safety, often exert a stronger predictive power on mode choice than built environment attributes alone. This suggests that, without addressing the latent demand for safety and social norms, infrastructural investments could lead to suboptimal modal shifts.

2.1.4 Limits of the 15-Minute City

While the concept of 15-minute city is often regarded as innovative and sustainable, a growing part of the literature warns about some of its possible negative effects. One of the major drawbacks is undoubtedly its physical determinism, in the sense of setting goals without specifying how or when they will be achieved (Khavarian-Garmsir et al, 2023). The proximity to services is automatically associated to sustainable behaviour, without further considerations on feasibility or structural barriers (Carvalho et al., 2025). Indeed, especially for cities that present the phenomenon of urban sprawl, the conversion to an x-minute city remains unfeasible without major infrastructural overhauls.

Another particular risk is eco-gentrification, or the process that leads 15-minute neighbourhoods to displace low-income residents by indirectly increasing property value (Elldér, 2024). Infrastructural interventions, such as pedestrianisation, improvement of cycling lanes or park installations, can lead to rising housing costs and to the displacement of lower-income households toward the suburbs. For this reason, it is important to address equity issues through adequate policies before the planning phase (Gould & Lewis, 2016).

These criticisms are particularly relevant for this thesis, since the available survey data (3.1) is rooted in peripherical areas. Understanding that travel behaviour is influenced also by the living location is crucial for the correct specification of the model.

2.1.5 Positioning of the Current Research: The Comparison between Non-Mandatory Trips and Commuting Trips

While established literature has extensively modelled urban transportation choices through Discrete Choice Models (Fale et al., 2025), the focus remains primarily on accessibility to essential services and commuting (HBW trips). In contrast, limited attention has been addressed to the impact of the 15-minute paradigm on non-mandatory trips, specifically leisure, visiting and shopping (Dong et al., 2025). These trips, being discretionary, are generally considered more flexible and context-dependent compared to rigid commuting patterns (Gim, 2017) and could, therefore, be more sensitive to latent variables related to urban quality. Indeed, mode choice studies find that people behave more flexibly in non-mandatory trips (Ortúzar & Willemsen, 2011) and are more sensitive to context than in commuting. Consequently, for these trip purposes, it is possible to explicitly link level of service variables (cost and time) and socio-demographic factors with personal habits (Loa & Habib, 2023) and built environment (Kalantari et al., 2025). This constitutes the first significant literature gap since investigating how mode choice determinants vary across discretionary trip purposes is crucial, especially those concerning psychological perceptions of the urban environment.

Furthermore, a second gap emerges regarding the role of these subjective factors within the 15-minute framework. Recent reviews on mode choice behaviour heavily emphasize that individual preferences, perceptions and latent attitudes are critical factors for travel behaviour (Ranjan & Sinha, 2025). However, despite the existing consensus in the transportation literature towards this topic, there is a clear lack of studies applying these psychometric constructs specifically to the 15-minute city paradigm.

This thesis aims to investigate if a positive latent attitude towards the 15-minute city concept effectively influences mode choice towards soft mobility, isolating the unobserved attitude effect from traditional socio-demographic characteristics. The difference across trip purposes will also be explicitly considered. It will be operationalized through a preliminary Multinomial Logit Model and, afterwards, via latent variables in a more complex hybrid ICLV (Integrated and Latent Variable) model.

2.2 Theoretical Framework of Discrete Choice Modelling

To proceed with the behavioural analysis of the 15-minute city, it is necessary to translate the qualitative decision-making process into a quantitative framework. The analysis of travel behaviour in this thesis relies on Discrete Choice Models, which describe the probability of an individual choosing a specific alternative from a finite set of mutually exclusive options.

Historically, transport planning relied on aggregate models (Four-Step Model) which analysed flows between zones rather than individual choices. However, since the 70s, the field shifted towards disaggregate behavioural models, which assume that the individual is the decision-making unit. As noted by various authors (Ben-Akiva & Lerman, 1985), this approach allows for a more accurate representation of policy impacts, as it accounts for the causal relationships driving individual behaviour rather than just observing general patterns at a zonal level.

2.2.1 Random Utility Theory (RUT)

The basic paradigm for these models is the concept of utility, that derives directly from microeconomics and the consumer theory. Developed in the 50's to relate goods purchased with characteristics of consumers, it introduces a set of important hypotheses:

- The different alternatives are described by quantitative attributes
- Choice set is fixed by dealer and is equal for all consumers
- All users have perfect information about the market
- Rational decision-making process, thus maximization of utility for every user.

This framework assumes that the decision-maker behaves as a rational actor (*homo economicus*), processing full information about the market to select the alternative that maximizes his personal utility (Train, 2009). However, unlike deterministic models, RUT acknowledges that the analyst cannot observe all factors influencing the choice.

Therefore, utility can be decomposed in two different elements. It is defined as:

$$U = V + \varepsilon$$

- V = Systematic utility, the observed component.
- ε = Error or disturbance, the unobserved component. Composed of measurement errors, taste variations in time and unobserved attributes in general.

Since real utility cannot be observed, the preliminary analysis will be made on systematic utility V, with i being one of the transport modes and j the number of explanatory variables.

$$V_i = \beta_0 + \sum_j \beta_j X_j$$

- β_0 = Intercept for the i mode (what is not explained by the j variables) or ASC = Alternative Specific Constant
- β_j = Coefficient of variable X_j

- X_j = Independent variable contributing to V_i

Therefore, in the easiest binary case the random user would choose the mode with highest utility. For this example, the modes "Motorized" and "Unmotorized" have been employed:

$$V_{unm} + \varepsilon_{unm} > V_{mot} + \varepsilon_{mot}$$

$$\varepsilon_{mot} - \varepsilon_{unm} < V_{unm} - V_{mot}$$

In order to proceed further, it is important to make some clarifications about the disturbance term. Assuming that every ε is Gumbel (Type I Extreme Value) distributed, an important property follows consequently: all errors are independent and identically distributed (IID). This implies that the difference $\varepsilon_{mot} - \varepsilon_{unm}$ follows a logistical distribution with zero mean: $\mu = 0$. Therefore, the term $V_{unm} - V_{mot}$ takes the same form.

2.2.2 The Logit Model

The Logit model is a discrete choice model based on the logistic regression, particularly useful when the endogenous alternatives are more than two (McFadden, 1977). In the simplest binary case, the probability of choosing one of the two alternatives changes with a logistic distribution law and depends on one variable.

For example, having "Motorized" and "Unmotorized" modes:

$$P(unm) = \frac{1}{1 + e^{-(c_{mot} - c_{unm})}} = \frac{1}{1 + e^{-(c^*)}}$$

- $P(unm)$ = Probability of choosing unmotorized mode
- c_{mot} = Generalized cost of motorized mode
- c_{unm} = Generalized cost of unmotorized mode

The higher the c^* (higher difference of cost towards motorized) the higher will be the probability for a random user to choose unmotorized transport mode.

As seen beforehand, since the systematic utility is logistically distributed, it can replace the generalized cost in the probability function. The previous equation becomes then:

$$P(unm) = P(V_{unm} > V_{mot}) = \frac{1}{1 + e^{(V_{mot} - V_{unm})}} = \frac{1}{1 + \frac{e^{V_{mot}}}{e^{V_{unm}}}} = \frac{e^{V_{unm}}}{e^{V_{mot}} + e^{V_{unm}}}$$

With this simplification the equation for multiple modes becomes evident:

$$P(i) = \frac{e^{V_i}}{e^{V_i} + \sum_j e^{V_j}}$$

j are the other modes considered besides i.

This formula highlights that the probability of choosing a mode increases as its systematic utility rises compared to the aggregate utility of all available options.

2.2.3 The Independence of Irrelevant Alternatives (IIA)

A direct structural consequence of the IID assumption in the MNL is the Independence of Irrelevant Alternatives (IIA) property. As defined by the first studies in the field (Luce, 1959), this property implies that the ratio of the choice probabilities of any two alternatives is entirely unaffected by the presence or characteristics of any other alternative in the choice set.

$$\frac{P(i)}{P(j)} = \frac{e^{V_i}}{e^{V_j}} = e^{V_i - V_j}$$

While this simplifies the overall estimation, it represents a significant limitation in transportation modelling. The IIA assumption fails when alternatives share unobserved attributes (as the famous "Red Bus/Blue Bus" paradox), potentially leading to biased predictions if the error terms are correlated.

Consequently, while the MNL serves as the robust baseline for this study, the validity of the IIA assumption must be rigorously tested using specific tools, such as the Hausman-McFadden test (Hausman & McFadden, 1984). This is to ensure that the grouping of transport modes, like the merge of walking and cycling, respects the independence requirement.

2.3 Advanced Modelling: Hybrid Choice Models

While traditional Multinomial Logit (MNL) frameworks provide a solid analytical baseline, they are structurally limited in capturing unobserved psychological attitudes. To address this issue, recent literature has highlighted the advantages of Integrated Choice and Latent Variable (ICLV) models (Vij & Walker, 2016).

This advanced technique refines the decision-making analysis by combining two methodological approaches:

- Traditional discrete choice models
- Structural equation models

The innovation compared to the previous framework is the SEM (Structural Equation Model) part, that contributes to the identification of the right latent variables and explains hidden behaviours not always captured by the exogenous variables of the logit model. It considers attitudes, perceptions and preferences of the general user, offering a more realistic picture of the decision-making process.

The complexity of the model stands in the various processes necessary to attain the utility function and therefore the mode choice. The regressors, usually socio-demographic or transportation related variables, are objective attributes of the user or of the alternatives. The latent variables are unobserved constructs linked to a set of observable indicators, such as Likert-scale survey questions, which should explain the underlying psychological attitude.

This dichotomy effectively illustrates the mathematical complexity of ICLV, operating simultaneously on two distinct levels.

- Observable Indicators: LVs (Latent Variables) are estimated using detectable manifestations of psychological patterns. They are not the cause of the attitude, but the consequence.
- Dual-Use Regressors: They act as predictors for both the utility functions and the latent variable equations.

Although ICLV architectures present a higher computational complexity and estimation challenges compared to a standard MNL, their superior ability to explicitly model complex behavioural dynamics makes them more appropriate for studying abstract concepts like the 15-minute city. As demonstrated by specific literature reviews (Bouscasse, 2018) and successful empirical applications in highly walkable environments like college towns (Kim & Li, 2023), the ICLV framework represents a powerful tool for incorporating complex behavioural theories and latent attitudes into mode choice modelling.

2.3.1 *Mathematical Formulation of the ICLV Model*

The utility of the ICLV model is comprehensive of both the objective regressors and the above defined latent variables:

$$V_{i,n} = ASC_i + \sum_j \beta_{i,j} \cdot X_{n,j} + \sum_l \gamma_{i,l} \cdot LV_{l,n}$$

- ASC_i: Alternative Specific Constant for mode i
- $\beta_{i,j}$: Estimated coefficient for variable j for alternative i
- $X_{n,j}$: Specific value of variable j for user n
- $\gamma_{i,l}$: Estimated coefficient for latent variable l for alternative i
- $LV_{l,n}$: Specific value of latent variable l for user n
- j are the explanatory variables

The latent variable formulation is given by:

$$LV_{l,n} = \beta_{0,l} + \sum_{j} \beta_{j,l} \cdot Z_{n,j} + \varepsilon_l$$

- $\beta_{0,l}$: Intercept of latent variable l
- $\beta_{j,l}$: Estimated coefficient for regressor j for latent variable l
- $Z_{n,j}$: Specific value of regressor j for user n
- ε_l: Structural disturbance term of latent variable l, $\varepsilon_l = \sigma_l \cdot \omega_l$
- σ_l: Standard deviation of error
- ω_l: Error term, distributed normally

The measurement part, consisting of indicators, is specified as follows:

$$I_{k,n,l} = \alpha_k + \lambda_k \cdot LV_{l,n} + \varepsilon_k$$

- α_k: Intercept of indicator k
- $I_{k,n,l}$: Indicator k of latent variable l for user n
- λ_k: Factor loading of latent variable l
- ε_k: Error term of indicator k

2.3.2 Advantages of the ICLV Models

Under certain conditions, the ICLV models offer unique advantages that reduced-form models cannot replicate (Vij & Walker, 2016). It is very important to identify these circumstances to understand whether the effort of producing a hybrid model is justified or not (Mariel et al., 2024). According to the previous author, if measurement indicators are available (survey responses) or the structural equations are particularly weak it could be beneficial to implement a structural section.

Moreover, the effect of an observable variable can be decomposed into direct and indirect effects through latent variables. For example, the influence of “Age” on the utility of a certain mode can be split into its indirect effect on "Attitude towards 15-minute city" and its residual direct effect on the choice. A standard model can only estimate the total cumulative effect but cannot distinguish the single mechanisms.

Since the available survey reports a lot of Likert-scale type questions as it will be introduced in section 3.1.3, using an ICLV model could reveal a structure that would not have been discovered without latent variables. This translates into the possibility of providing policy insights regarding attitudes that a standard MNL model could not deliver.

Various studies report that latent variables predict mode choice more accurately than objective variables alone (Chen & Li, 2017; Jameel & Abdulhussein, 2024; Lu & Gan, 2024), such as, for example, travel time (Kamargianni & Polydoropoulou, 2013). In particular, the environmental dimension (Toorzani & Rassafi, 2022) has been found to be highly significant for travel behaviour and is remarkably relevant for the 15-minute city concept analysed in this work. Furthermore, other findings identified the latent variables concerning convenience, infrastructure and risk perception as three of the most determinant factors for bicycle use (Fernández-Heredia et al., 2014). Similarly, in recent studies, the researchers (Adamidis et al., 2025) incorporated pro-car and pro-environment latent constructs within a HCM (Hybrid Choice Model) framework, highlighting the relevance of attitudinal dimensions in influencing transport mode preferences and mobility policies.

These studies confirm that omitting psychological factors leads to an underestimation of the potential for modal shift. By explicitly modelling the "Attitude towards the 15-minute city" and " Perceived Easiness of Car Use," this thesis adopts the framework of established literature, extending the analysis to the specific context of proximity-based urban planning.

2.4 Determinants of the Mode Choice and Trip Purpose

The specification of a discrete choice model requires a rigorous selection of explanatory variables. According to the literature, mode choice is rarely the result of a single factor but rather a complex combination between individual socio-demographic characteristics, spatial attributes, and psychosocial perceptions. This section reviews the principal variables identified in previous studies that will be tested in the context of the 15-minute city. Methodologically, it should be noted that the majority of these reviewed works derive their behavioural findings from standard discrete choice frameworks, primarily Multinomial Logit models. This indirectly reveals that ICLV models remain less common compared to standard discrete choice approaches, which justifies their use in the present research.

2.4.1 Socio-Demographic Factors

Individual characteristics are often the strongest predictors of travel behaviour, defining the socio-economic situation of the user.

- Gender: Travel patterns exhibit significant heterogeneity across genders. Diverse authors investigated about transit from the women perspective (Loukaitou-Sideris, 2014), suggesting that men and women experience the built environment differently due to a distinct safety perception. Moreover, evidence shows a difference in travel needs. Women often chain multiple short trips (tours) due to household constraints (Hanson & Pratt, 1995), which increases their sensitivity to local accessibility and safety. Furthermore, recent studies in the post-pandemic context (Murooka et al., 2025) indicate that gender remains a significant differentiator in the adoption of sustainable modes, with women showing a higher aversion to perceived risks in public spaces but also being more prone to use “greener” modes.

- Age: The relationship between age and public transport use is typically treated as non-linear. Indeed, literature highlights that mobility needs are different across life stages (Tyrinopoulos & Antoniou, 2013). Young adults (captive users) and the elderly often show higher utility for public transport and walking due to economic/health constraints or incentives, while active adults are more prone to private car usage due to time pressure and higher motorization rates. This suggests a "U-shaped" utility curve rather than a linear progression, with a peak of motorized use in the period between 40 and 50 years old.

- Education and Income: Economic theory establishes a strong correlation between education level and income accumulation (Mincer, 1974). In transport surveys, income data is frequently subject to a non-response bias, especially for richer groups, creating incongruence in the results if the issue is not correctly addressed. Education, on the other hand, is an information that is easier to acquire. Furthermore, income is generally more fluctuating between periods, while education is considered to be a more stable indicator of long-term economic capability. Consequently, various authors argue that education level acts as a robust proxy for socio-economic status (Galobardes et al., 2006). For this reason, using the two at the same time could lead to problems of collinearity. Generally, higher education levels correlate with higher income, which historically increased the probability of car ownership. However, recent trends suggest that highly educated individuals in dense urban cores are increasingly shifting towards multimodal behaviours (George et al., 2025). These assumptions will be verified with the available dataset.

2.4.2 Built Environment and Psychosocial Factors

Beyond personal traits, the physical attributes of the urban landscape and the user psychological response to them are critical.

- Density and Crowding: The impact of density on mode choice presents a dichotomy. While high density reduces travel distances, promoting active travel (Colaço & de Abreu e Silva, 2025; Frank et al., 2000), it can also lead to the perception of crowding (Tyrinopoulos & Antoniou, 2013). Thus, while physical density supports public transport viability, the perception of crowding can act as a deterrent, negatively affecting the utility of mass transit and pushing users towards private modes. For this reason, its ambiguous role will be investigated furtherly in the models.

- Accessibility and Proximity: Central to the 15-minute city concept is the proximity to amenities. Various authors demonstrate that transport accessibility and proximity, that is defined by the ease of reaching different services, is a primary driver for unmotorized trips (Fernández Núñez et al., 2024). However, land use alone has a "limited effect" if not coupled with a favourable perception of the environment or other factors (Murooka et al., 2025).

- Car Ownership and Mobility Assets: The possession of mobility tools (cars, transit passes) is often the dominant predictor in mode choice. Car ownership is frequently defined not just as a variable but as a hierarchical decision that conditions daily choices (Ben-Akiva & Lerman, 1985). The “saturation” effect, characterized by the diminishing marginal utility of additional vehicles (Dargay, 2001) will also be examined and explicitly incorporated into the modelling strategy.

2.4.3 Trip Purpose Characterisation

Trip purpose plays a central role in how individuals respond to the 15-minute city context. Several studies highlight its correlation with mode preferences (Fang et al., 2014; Kim et al., 2020; Scheepers et al., 2014). While commuting trips are typically constrained by fixed schedules and specific destinations (workplaces or education spaces); non-mandatory trips, such as shopping, leisure, and social activities, are generally more flexible in both timing and destination choice

(Upahita et al., 2022). This flexibility signifies that discretionary trips are particularly sensitive to local accessibility and neighbourhood-scale amenities, which are at the core of 15-minute-city strategies.

In this environment, the proximity and diversity of services within walking or cycling distance can substantially influence the modal choice of non-mandatory trips (Negm et al., 2023). For commuting trips, structural constraints such as job location, working hours and employer duties often limit the potential for an unrestricted modal behaviour, thus improvements in local accessibility could have a more modest effect on mode choice compared to non-mandatory travel (Mosquera et al., 2023). Moreover, psychosocial factors (perceived safety, comfort, neighbourhood pride, perception of congestion and attitudes toward soft mobility) can influence trip purpose as well (Bhagat-Conway et al., 2022).

Recognising these differences is essential when modelling mode choice for the 15-minute city perspective. It implies that the same infrastructural intervention can have different impacts across trip purposes, increasing active and public transport use for non-mandatory trips while only partially affecting commuting patterns. Consequently, distinguishing commuting and non-mandatory trips is required to have a satisfactory modelling framework for mode choice in the 15-minute city.

2.5 Synthesis of the State of the Art and Research Hypothesis

The review of the existing literature presented in this chapter highlights the evolution of the mode choice models over the years, moving from aggregate trip-based approaches to disaggregate behavioural models (Ortúzar & Willemsen, 2011; Ben-Akiva & Lerman, 1985). As identified in section 2.1.5, applying these frameworks to the 15-minute city paradigm reveals two different literature gaps: the predominant focus on commuting rather than non-mandatory trips and the limited investigation on psychological factors. The structural dimension of the latest models has proved to be fundamental to address the hidden behaviours of the population (Johannson et al., 2006), especially in the context of the 15-minute city. In proximity-based settings, understanding mode choice requires capturing not only spatial accessibility but also residents' perceptions, habits and overall preferences. The determinants of utility, under the circumstances of the present case study, cannot only depend on simple observable sociodemographic characteristics (Vij et al., 2013).

This thesis aims to explore the difference in behaviour among commuting trips and non-mandatory trips from the 15-minute city point of view. Thanks to the methodological insights discovered in section 2.4, the comparison between the simple MNL (Multinomial Logit, 3.4.1) model and a more complex ICLV (Integrated Choice and Latent Variable, 3.4.2) model will identify the importance of psychosocial constructs in the determination of utility for the different modes. This will allow to identify if their significance differs by trip purpose, or, in other words, if adopting a hybrid choice model is useful only for certain types of travels.

In conclusion, the State of the Art confirms the possibility of adopting a Hybrid Choice Model to fully capture the complexity of urban mobility. The following chapter 3 details the methodological approach designed to test these hypotheses empirically.

3. Case Study and Methodology

3.1 Case Study: Survey Implementation and Related Dataset

The survey originating the data used in this study was conducted in eight localities in the municipality of "Vila Franca de Xira", a suburban area in the northern part of Lisbon Metropolitan area. Available data was gathered in a confined area of a diameter of 20 km. Given the fact that the study area encompasses different urbanized cores, the spatial variability in the survey data will be furtherly investigated afterwards (3.2.4). A random quota-based intercept sampling process was used to gather the data, in order to get a representative cross section of the population. Specifically, the data collection campaign was conducted in May 2025 across the target locations and was structured as a face-to-face intercept interview, where participants were approached in public spaces (squares, parks and transit hubs).

3.1.1 Geographic Characterisation of the Study Area

The study focuses on the northern-eastern part of the Lisbon Metropolitan Area (AML), specifically targeting the Vila Franca de Xira municipality (including the homonymous parish). This territory was selected due to its strategic role as a primary axis connecting the peripherical residential districts to the central part of Lisbon. The area is characterized by a high demographic density and a strong dependence on the northern railway line (Linha do Norte) and the A1 motorway. Specifically, the metropolitan rail service is provided by a portion of the abovementioned railway line, called "Azambuja Line", serving only the outer suburbs of the Portuguese capital.

The survey covered eight distinct administrative units, called "freguesias" in Portuguese, namely "parishes". After an administrative reorganization undergone in Portugal in 2013, many parishes have been merged into unions of parishes (União de Freguesias, U.F.), thus the official data from the latest census (2021) are only available in this format, where the study area is split in six zones. Key information on these units is reported in Table 1.

Table 1: Demographic and Geographic Characterization of the Study Area (Source: own elaboration based on data of the 2021 Portuguese census)

Location	**Population (inhabitants)**	**Area (km^2)**	**Density (inhabitants/(km^2)**
Alhandra, São João dos Montes e Calhandriz	12866	27.54	467
Alverca do Ribatejo e Sobralinho	36120	23.92	1510
Castanheira do Ribatejo e Cachoeiras	8266	26.78	309
Póvoa de Santa Iria e Forte da Casa	40905	9.16	4466
Vialonga	21261	17.52	1214
Vila Franca de Xira	18336	212.86	86
Total	*137754*	*317.78*	*433*

As it is possible to perceive from the data, the densest cluster is by far the "Povoa de Santa Iria e Forte da Casa" U.F.. It is characterized by a widespread vertical residential development and the proximity to the city centre. The higher the distance from the downtown, the lower the density becomes, with "Castanheira do Ribatejo" e "Vila Franca de Xira" exhibiting the characteristics of low-density rural settlements.

Nevertheless, the mobility context is similar in all the study area, with the constraints of the Tagus River to the east and steep topography to the west, naturally creating a linear corridor characterized by a certain uniformity in transport services. Figure 1 illustrates the spatial distribution of the respondents' places of residence across the whole study area, which are marked with red dots. The railway line and the A1 motorway act as a connection between the residential areas and the central area of Lisbon.

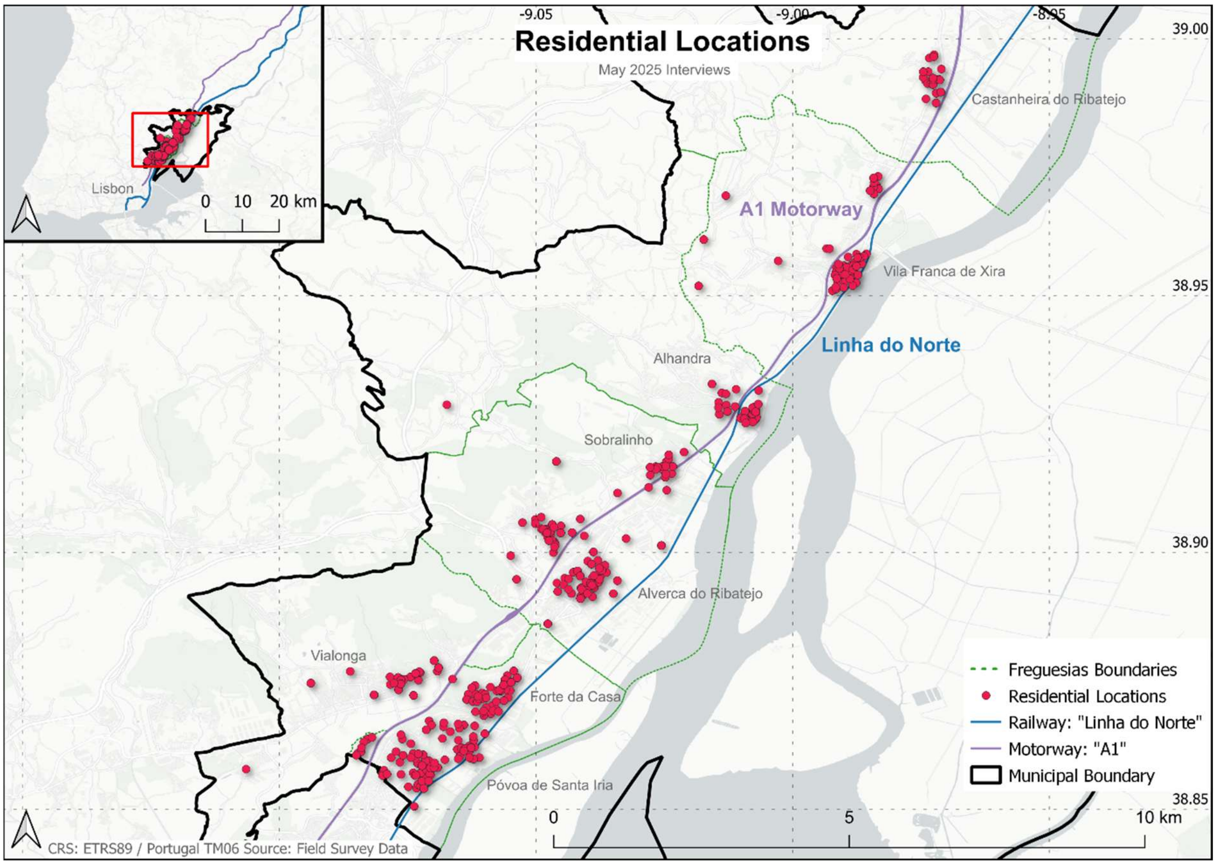

Figure 1: Residential Locations across the Study Area with Main Transportation Axes (Source: own elaboration using QGIS based on survey data)

3.1.2 Data Collection and Sampling Plan

To guarantee the statistical representativeness of the sample and mitigate the selection bias present in random street intercept methods, a "Quota Sampling Strategy" was employed. This sampling technique involved the definition of specific targets based on gender and age segments, representing the demographic distribution of the selected locations. Field researchers were instructed to casually approach individuals on the street until the pre-defined quotas for each demographic group were filled. This approach ensured that the final dataset closely approximates the population structure, avoiding the over-representation of specific groups (students or elders) that typically occurs in random street surveys. Although the map reports the residential location of the respondents, interviews were conducted in shared urban spaces within the selected localities.

The survey questionnaire was planned by two students from "Instituto Superior Técnico" with the supervision of Professor Joao de Abreu e Silva. The questions were formulated by his research team after reviewing the literature on the concept of the 15-minute city and focused primarily on neighbourhood satisfaction, built environment assessment and habitual travel patterns. The practical implementation was carried out by a contracted company, which provided a team of trained interviewers to administer the questionnaires on-site, through the CAPI (Computer-Assisted Personal Interview) method. The whole process was funded by the FORTHCOMING (FOsteRing THe City Of proximity through Maas INteGration) project[1]. It aims to adapt central city strategies to suburban context to improve liveability and sustainability of the peripheral areas.

The gathered sample is composed of over one-thousand observations, precisely 1055. Since it is mostly framed by ordinal and categorical variables, to statistically determine the adequacy of the sample size the following formula is used:

$$n = \frac{p(1-p)z^2}{e^2}$$

- n = minimum sample size
- p = the expected proportion of the population with the characteristic of interest (since p is unknown the value of 0.5 is used to maximize n and be more prudent)
- z = the z-score from the normal standard distribution to reach the desired confidence level
- e = the maximum allowable margin of error

The aim is an error smaller than 5%. Using p = 0.5, z = 1.96 to have a 95% confidence interval and reverting the equation to find the error term, the result is $e \cong 0.03$. The sample size is perfectly adequate for the needs of this model.

3.1.3 Survey Contents and Resulting Database

The survey consists of three distinct and specific sections:

- Socio-demographic profile of the user: This section includes data on the respondent personal background, including age, gender, education level, household size, etc… These variables are the basic regressors for the discrete choice model.
- Mobility assets: Collects information regarding mobility tools of the user, such as number of vehicles, license, transport card, etc…

[1] Official project website and documentation available at: https://eurocities.eu/projects/forthcoming/

- Attitudinal assessment: Consists in a series of statements relative to the 15-minute city concept and the living neighbourhood. Respondents evaluated, principally on a Likert-scale, features like safety perception, satisfaction with local amenities, traffic perception, etc… This part enables the generation of an ICLV (Integrated and Latent Variable) model and its structural part.

The whole database description, including extensive explanation of each variable, is presented in Appendix A1.

3.2 Data Processing and Validation

3.2.1 Identification of the Model Dependent Variable

To understand how the database is structured it is important to perform univariate analyses, especially for the endogenous variable. Identifying the strengths and weaknesses of the survey is the first fundamental step toward model specification. Consequently, it is necessary to examine preliminarily the endogenous variable. Indeed, having a problem with it could compromise the whole model and, thus, the necessary countermeasures must be taken as soon as possible.

The primary dependent variable in this study is extrapolated from the following survey question: "Last week, what was the mode of transport you used the most for the following trip purpose?" Therefore, the analysis does not rely on individual trip observations, but on the respondents self-reported predominant travel mode for each trip purpose during the previous week. To avoid the common "aggregation bias" in more general mobility studies (Jeong et al., 2022), where commuting trips are mixed with leisure ones, the survey questions concerning mode choice were segmented by trip purpose, considering five distinct categories (with their respective codes indicated in brackets):

- Commuting (Mode_commuting_Opt): This category represents the mandatory trips to work or school. It is characterized by high rigidity and frequency, usually occurring on a daily basis. In this case, the choice of the mode is mainly driven by travel time and reliability, meaning that the user is generally less sensitive to the urban environment compared to other purposes.
- Visiting (Mode_visiting_Opt): This purpose includes trips made to visit friends or relatives. Unlike commuting, these trips are non-systematic and often happen during off-peak hours or weekends. The decision-making process here is less dependent on minimizing time and more influenced by convenience and social habits.

- Leisure (Mode_leisure_Opt): Recreational trips, such as going to the park or practicing sports, represent the most flexible type of mobility. Since the activity is discretionary, the choice of the mode is highly sensitive to the quality of the urban space and the attitude of the user. A similar pattern to that of visiting trips is expected.
- Shopping (Mode_shopping_Opt): They are considered "maintenance" activities. Even if they are essential, they have more flexibility than commuting. For this purpose, the "cargo constraint" is very important, because the need to carry goods often pushes the user to choose the private car, regardless of the distance or the proximity of the service.
- 15-minute area (Mode_15-min_Opt): Finally, the survey includes a specific question about the mobility within the residential area. It asks: "How do you usually travel in the area where you reside (within a radius of 800 meters)?". This variable is different from the others because it acts as a spatial filter. It isolates the behaviour in the local scale, showing whether the user relies on the car even for short distances due to barriers and habits or if he uses active modes in his neighbourhood instead. Therefore, this category is not mutually excluding the previous four.

Regarding the modal alternatives of the endogenous variable, the modes identified in the survey were 7 in total, but it is worth noting that a number of participants did not answer this question. Table 2 provides a general overview that depicts the frequencies of each mode.

Table 2: Absolute and Relative Frequencies (Total Answers vs. Total Sample) of Transport Modes per Trip Purpose (Source: own elaboration based on survey data)

Transport Mode	**Commuting**	**Visiting**	**Leisure**	**Shopping**	**15-min**
Car	444 (42.1%; 63.4%)	392 (37.2%; 44.4%)	253 (24.0%; 26.7%)	530 (50.2%; 53.3%)	262 (24.8%; 24.9%)
Walking	54 (5.1%; 7.7%)	390 (37.0%; 44.2%)	622 (59.0%; 65.4%)	391 (37.1%; 39.3%)	709 (67.2%; 67.3%)
Public Transport	186 (17.6%; 26.6%)	88 (8.3%; 10.0%)	61 (5.8%; 6.4%)	67 (6.4%; 6.7%)	59 (5.6%; 5.6%)
Bicycle	1 (0.1%; 0.1%)	6 (0.6%; 0.7%)	10 (0.9%; 1.1%)	3 (0.3%; 0.3%)	18 (1.7%; 1.7%)
Taxi	3 (0.3%; 0.4%)	4 (0.4%; 0.5%)	2 (0.2%; 0.2%)	3 (0.3%; 0.3%)	1 (0.1%; 0.1%)
Motorcycle	12 (1.1%; 1.7%)	3 (0.3%; 0.3%)	2 (0.2%; 0.2%)	1 (0.1%; 0.1%)	4 (0.4%; 0.4%)
Total Answers	*700*	*883*	*950*	*995*	*1053*
N.A.	355 (33.7%; -%)	172 (16.3%; -%)	105 (10.0%; -%)	60 (5.7%; -%)	2 (0.2%; -%)
Total Sample	*1055*	*1055*	*1055*	*1055*	*1055*

At first glance, it is possible to observe that, besides "N.A.", the two predominant modes for all trip purposes are "Walking" and "Car". Although their percentage varies considerably between the different travel categories, a pattern is recognizable. For obvious reasons, commuting trips exhibit the lowest frequency of walking trips, while for discretionary travel their number increases significantly. 15-minute and leisure are the purposes for which unmotorized mobility reaches its peak, whereas for commuting and shopping a predilection for car is observed, for the reasons stated in the 2.4.3 section. Visiting trips are more nuanced, since the number of car users is roughly equivalent to walking people.

On the other hand, "Bicycle", "Motorcycle" and "Taxi" modes are negligible for all trip purposes and, therefore, will be jointly considered with other modes, given their low frequency. Indeed, a minimum number of observations is required to properly proceed with the framework. The only other significant category is "Public Transport", with more than 50 observations for each purpose. The number of observations is similar across discretionary trips, while it increases by a significant amount for commuters. Due to the aforementioned considerations, a model with an endogenous variable that has 3 different options seems the most optimal choice.

It is important to note that commuting trip purpose exhibits the highest frequency of “N.A.” responses. This result is consistent with the demographic composition of the sample and most likely reflects retired individuals. While these respondents don’t perform work-related trips anymore, they continue to participate in discretionary activities such as shopping, leisure and visiting, explaining the considerably lower number of missing responses for the remaining trip purposes.

3.2.1.1 Leisure Trip Purpose Exclusion

Although leisure trips were initially categorized as a distinct purpose in the survey, preliminary analyses revealed that the travel behaviour for recreational scopes was highly consistent with the visiting one. To analytically evaluate the similarity among non-mandatory travel patterns, a Multi Correspondence Analysis (MCA) was conducted on the discretionary mode choice variables (Visiting, Leisure and Shopping).

MCA is an advanced multivariate statistical technique designed to detect associations between categorical data. In the present case, the analysis was performed using SPSS, considering the mode choices stated for the three discretionary trip purpose variables: “Mode_visiting_Opt”, “Mode_leisure_Opt” and “Mode_shopping_Opt”. The variables are labelled as “Opt” since they were recoded to include only three main alternatives: “Motorized”, “Unmotorized” and “Public Transport”. The methodological justification for this segmentation is detailed in the subsequent section 3.2.2.

The objective of the analysis is to project the three variables in a two-dimensional plane to measure their relative variance. In the geometric representation summarized in Figure 2, the coordinates of each variable indicate the correlation of the latter with the two principal dimensions extracted by the model. Indeed, a similar amount of variance explained for each component translates into a higher association between trip purposes.

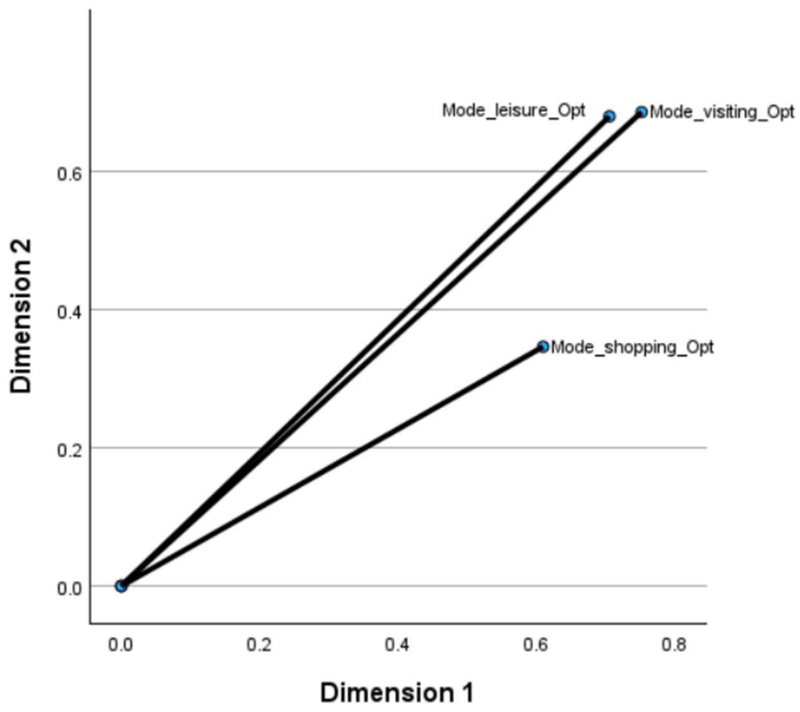

Figure 2: Discrimination Measures for Discretionary Trip Purposes (source: own elaboration through SPSS)

As illustrated, the vectors representing leisure and visiting choices are almost perfectly overlapping across the principal dimensions, demonstrating a similar underlying behavioural structure. This indicates that respondents tend to adopt a similar modal strategy for both recreational activities and social interactions. Conversely, shopping exhibits a distinct trajectory, confirming its unique dynamics, such as the "cargo constraint" explained beforehand. Despite the difference from a mandatory trip like commuting, it is labelled as a "maintenance" activity, which is notably influenced by specific spatial and physical boundaries.

Consequently, maintaining both the visiting and leisure models would lead to analytical redundancy, increasing computational complexity without improving explanatory power. For this reason, the leisure model was excluded from the final estimation, assuming the visiting model is a robust indicator for both social and recreational travel.

3.2.2 Data Cleaning

Two data cleaning and recoding actions were performed:

- Merging of categories in the dependent variable to improve the significance of the multinomial model

- Removal of the “N.A.” observations

Regarding the first procedure, the observed modes have been compacted into three different alternatives:

- Motorized (Reference category) = Car + Motorcycle + Taxi
- Unmotorized = Walk + Bike
- Public Transport

This choice was pursued to reduce the variability of the data and to avoid a high number of empty cells in the frequency matrix (pivot table). Indeed, the categories with a low number of observations have a worse modelling precision and thus it’s important to have relatively balanced classes. The objective is to merge categories that share a similar travel behaviour. Therefore, the proposed segmentation aims to maximize inter-class heterogeneity while minimizing intra-class variability, making sure that each group represents a distinct behavioural pattern.

In this study, Taxi and Ride-hailing services have been classified within the Motorized nest rather than Public Transport. This classification is justified by the operational characteristics of the mode: like private cars, taxis offer on-demand, door-to-door service that provides privacy and flexibility, unlike the fixed-schedule and fixed-route nature of public transport. Furthermore, from a utility modelling perspective, taxis are subject to the same road congestion constraints as private vehicles, making them sensitive to the “Perceived Easiness of Car Use” latent variable (LV_2) analysed afterwards (3.4.2.3), in the same way as personal cars.

After the construction of the model, the expectation is to observe a difference in travel behaviour for non-mandatory trips (leisure, visiting, shopping) compared to mandatory ones (commuting) and in related determinants. This can prove the positive effects of the 15-minute city on the sustainable mobility modes of its residents. The proximity to essential amenities should have a significant impact on non-mandatory trips, while being less influent for commuting, which is primarily constrained by the fixed location of the workplace.

The second part of the work consisted in cleaning the available data. In fact, some respondents did not answer all the questions thus leaving some vital cells (like transport mean) empty. Since the econometric models used in this work cannot use observations with a missing endogenous variable, there are two viable options to proceed further:

- Listwise deletion: erasing observation with a missing endogenous variable, reducing the sample size.

- Imputation of the missing values of the dependent variable: generally considered inappropriate because it distorts the statistical meaning of the model.

The main reason to discard the second option is that imputation must be done with variables that later will be used for modelling, thus artificially modifying the data that should be explained afterwards. In short, this approach:

- Uses predictors to estimate the missing values and then uses them again to run the model: circular reasoning fallacy.
- Reduces the variance: doing a regression creates Y values closer to the mean increasing artificially statistical significance and reducing standard deviation.
- Loses the real-world representativeness: the framework is not based anymore on real observations, which is essential for regression models.

Indeed, the dependent variable has to be observed and not estimated.

Regarding the independent ones it's radically different since the abovementioned problem does not exist. Nevertheless, the exclusion of certain observations could lead to a bias in the model. Particularly concerning "Income Range", often the wealthiest people tend to answer this question less than other income classes. Therefore, the elimination of these observations could not reflect the true statistical significance of the explanatory variable. For this reason, it's important to provide further analysis and inspect if the missing data are randomly distributed.

The most useful tool is the comparison of the means. After the creation of a dummy variable that has "0" for missing value and "1" otherwise, it's possible to proceed with the Mann-Whitney non-parametric test. It allows to evidence if there are significant changes between the means of binary values in relation to other independent variables. Moreover, it produces an output that gives the significance value of this "shift", to understand if the values are MCAR = Missing Completely At Random or MNAR = Missing Not At Random. This issue will be investigated in the appropriate section 4.1.1.1.

3.2.3 Recoding Independent Variables

Another issue before proceeding towards model specification is the management of ordinal explanatory variables. The safe option is to treat them as factors (not continuous), even if the model becomes heavier. Indeed, creating dummies consumes degrees of freedom, but since the sample size is large it should not be a problem. Moreover, this allows, as explained above, for non-linear effects, especially when the categories are not numerically defined, like with "High", "Medium" and "Low". Indeed, it is very complicated to evaluate if the difference between each category is constant. It might be that moving from "Low" to "Medium" has a large impact on utility but moving from "Medium" to "High" has almost a null effect.

3.2.3.1 Recoding Non-Metric Variables

To mitigate the negative effects mentioned previously, the recoding of many ordinal variables has been made, reducing the categories to 3 or 2 for each one. Particular emphasis has been placed on the selection of the right classes, with the objective of decreasing intraclass variance and increasing interclass one. The recoding process has been standardized for most variables and embraces socio-demographic features ("Income_Range" into "Income_Range_Ordinal", "Education_Level" into "Education_Level_Ordinal", etc), spatial and trip attributes ("Density" into "Density_Opt", "Traffic" into "Traffic_Opt", "Avg_Trip_Time" in "Avg_Trip_Time_Ordinal", etc) and selected psychometric indicators used in the latent constructs (see A1: Database Codebook). During the data processing phase, it was noticed that, especially for the ones that were originally measured on a 5-point Likert scale, from "Strongly Disagree" to "Strongly Agree", the distribution of responses was not uniform and some categories (especially the extreme negative option) contained less than 5% of the observations. This creates an imbalance that can impact the robustness of the model. Indeed, using a large number of categories could reduce model convergence and increase the risk of multicollinearity among the resulting dummy variables.

For this reason, as explained beforehand, the Likert variables were recoded into new variables with fewer categories. This choice was performed to improve the general reliability and statistical impact of the factors in the MNL and ICLV models. The two recoding options are presented in Table 3.

Table 3: Recoding of Likert-Scale Variables (Source: own elaboration)

Original Likert Value (5-cat)	Recoding 1 (3-cat)
5 - Strongly agree	2 – Agree
4 - Somewhat agree	
3 - Neither agree nor disagree	1 - Neither agree nor disagree
2 - Somewhat disagree	0 – Disagree
1 - Strongly disagree	

The recoding into two categories, creating de facto binary dummies, was avoided in the majority of the circumstances. In fact, deciding if the neutral class should correspond to the positive or negative grouping can appear as a subjective choice. In these specific and rare cases, the upper categories (4 and 5) were aggregated to represent a definitive positive perception or high satisfaction, coded as "1". Conversely, the neutral and negative categories (1 to 3) were merged into a single baseline group, coded as "0". This binary recoding is performed primarily for the reduction of computational complexity.

3.2.3.2 Recoding the "Age" variable

For continuous variables, the above approach is not applicable and usually not necessary. Indeed, the computational power required to estimate the coefficients of a metric factor is lower, which favours model parsimony. It should be noted that this removes the possibility of capturing non-linear effects of a certain socio-demographic attribute. It is exactly the case of the "Age" variable, which has been treated differently for the reasons stated in 2.4.1. Initially scalar, it has been decided to recode it into two different formats:

- Ordinal variable with three classes, as young people, active adults and elders behave differently regarding transport modes. If it was a linear scalar variable, it could not represent the different line of conducts of the three age brackets listed above. Moreover, the literature confirms that it is the most standardized way to address the issue.
- Scalar variable with quadratic term, in the form of $U = \beta_1 \cdot Age + \beta_2 \cdot Age^2$. In this way, the nonlinear behaviour between youngsters and seniors could be depicted with a U-shape function.

Regarding the ordinal variable coding, the analysis to find the best age brackets was performed in the SPSS software, with a tool called "CHAID" (Chi-Square Automatic Interaction Detection). In the form of a decision tree, it determines the most statistically significant segmentation of a dataset based on the relationship between a dependent variable and a set of independent predictors. Executing multiple iterations, it tests various cut-points to merge categories that are statistically homogenous with respect to the dependent variable. In this way, it evaluates the best possible splits, as it is possible to appreciate in Figure 3.

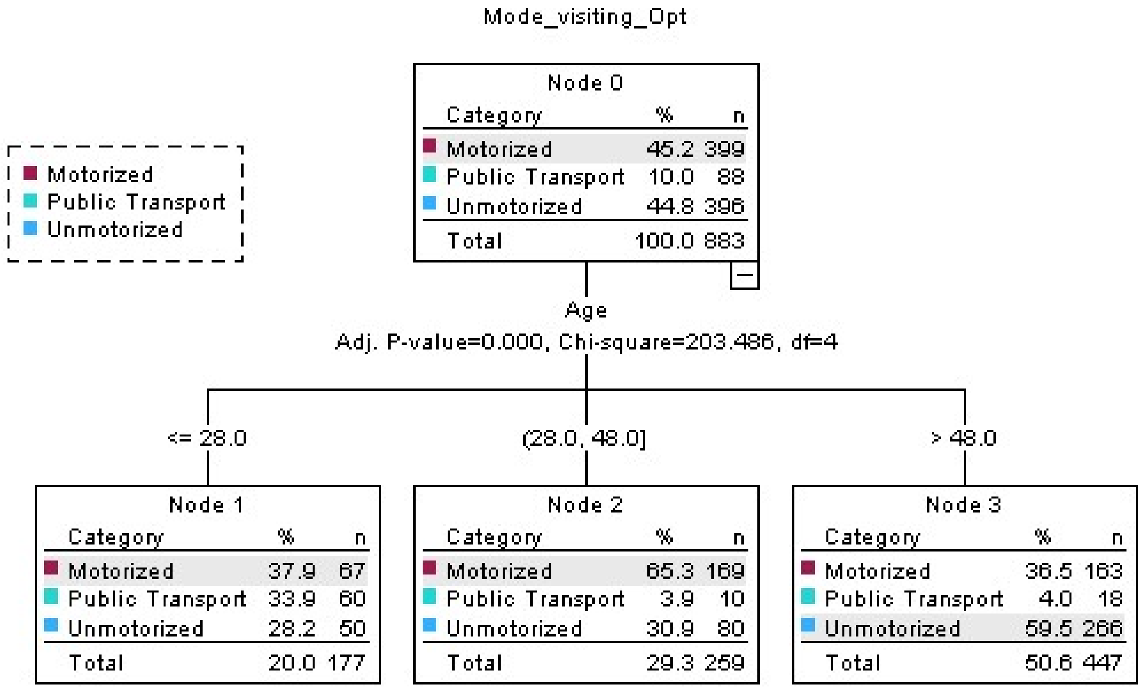

Figure 3: CHAID Output for "Age" Variable (Source: own elaboration using SPSS)

The reading of the results has to be critical, considering common sense but also the previous literature findings. The first bracket seems to indicate the group of young people mentioned before, that rely heavily on public transportation, while the second and third ones are less clear. Especially the latter, that is composed by people in very different stages of life. At this point there are two possible choices:

- Data-driven variable: keeping the CHAID results, that are the most significant as possible. This does not reflect the standardized classification but improves the overall fitness of the MNL model.
- Theory-driven variable: using the most conventional age brackets to separate life cycles, thus:
  - Young people: $Age < 30$.
  - Active adults: $30 \leq Age < 65$.
  - Elders: $Age \geq 65$.

The results obtained in the following part (4) will determine which of the two classifications is the most appropriate for the compromise between model significance and theoretical framework.

3.2.4 Dealing with Location Biases

Before proceeding with the validation of the whole model, it is important to investigate the possible presence of a location bias in the data. The survey was gathered in 8 different locations of the Lisbon Metropolitan Area (3.1.1), that were originally represented by a string variable ("Local") in the survey (see A1 for details). Please note that this categorical variable is not representing where the respondents live or perform any specific activity, although it is somewhat expected that a correlation between location of activities and location of the survey exists. A first analysis can be done through a crosstabulation between the mode used and the gathering place. In Table 4 it is noticeable that for some localities the distribution percentages are significantly different.

Table 4: Modal Distribution by Location (Source: own elaboration based on survey data)

Location	Motorized	Public Transport	Unmotorized	Total (N)
Castanheira do Ribatejo	20 (36.4%)	10 (18.2%)	25 (45.5%)	55
Vila Franca de Xira	72 (56.3%)	16 (12.5%)	40 (31.3%)	128
Alhandra	32 (43.2%)	5 (6.8%)	37 (50.0%)	74
Vialonga	43 (50.0%)	11 (12.8%)	32 (37.2%)	86
Póvoa de Santa Iria	61 (29.0%)	9 (4.3%)	140 (66.7%)	210
Forte da Casa	52 (59.1%)	5 (5.7%)	31 (35.2%)	88
Sobralinho	15 (42.9%)	9 (25.7%)	11 (31.4%)	35
Alverca do Ribatejo	104 (50.2%)	23 (11.1%)	80 (38.6%)	207
Total	*399 (45.2%)*	*88 (10.0%)*	*396 (44.8%)*	*883*

Indeed, it is possible to observe two evident outliers highlighted by the empirical data. The observations of unmotorized mode users in Póvoa de Santa Iria and public transportation users in Sobralinho are well above the respective average.

- Póvoa de Santa Iria: This location behaves like a strong outlier in terms of active mobility. With a 66.7% share of unmotorized trips, it vastly exceeds the overall average of 44.8% and records the highest level of active transport in the entire dataset. On the contrary, its reliance on motorized vehicles (29.0%) and public transport (4.3%) is significantly lower than the municipality average, despite having the largest number of survey respondents (N=210).
- Sobralinho: On the other hand, this second outlier exhibits a clear deviation in public transport usage. While most parishes stand below the 15% mark, Sobralinho registers a

remarkable 25.7% share for public transport, more than double the general average. However, when interpreting this specific outlier, it is crucial to account for the sample size since, with only 35 observations, Sobralinho represents the smallest group in the dataset, making its percentage distribution more susceptible to statistical errors. Therefore, the significance of the eventual dummy variable will be commented successively (4.1.1.2).

Furthermore, as engineers, it is fundamental to investigate the practical reasons of both cases. Since the difference between the two outliers and the other places seems to be pronounced, there should be some unique feature that characterize the mode behaviour of its inhabitants.

Concerning Póvoa de Santa Iria, this can be explained in various ways:

- Quality of pedestrian infrastructure. The presence of the "Parque Linear Ribeirinho Estuario do Tejo" critically differentiate the area compared to other locations. This infrastructure provides a safe pedestrian corridor that spans from the urban core to the train station and the riverfront. Transforming walking from a necessary activity to a pleasant activity enhances the enjoyability of active mobility.
- Residential area / Dormitory suburb. Unlike the more industrial layouts of most neighbouring parishes, Povoa de Santa Iria is characterized by a high-density housing area, aligning perfectly with the 15-minute city policies. The reduced distances between amenities increase the attractivity of walking or biking.
- Topographic and geographical advantage. Povoa de Santa Iria is well connected to the Carris (the main bus transit operator in Lisbon Metropolitan Area) suburban network and to the Azambuja trainline. Moreover, the flat terrain lowers the physical barrier for soft mobility.

Regarding Sobralinho, the public transportation shift can be explained by a few unique features of the location:

- Adverse motorway accessibility: Unlike the major nodes of Alverca or Vila Franca de Xira, which benefit from direct interchanges with the A1 Motorway, Sobralinho road connectivity is structurally dependent on the N10 National Road. This linear corridor exhibits frequently high levels of traffic, imposing high generalized costs on private vehicle use, due to congestion, traffic signals and lower speeds. The rail service, through the Azambuja line, acquires a significant comparative advantage in this situation.
- Station oriented morphology: Sobralinho exhibits a unique topographical layout across the locations in the sample: it is a high-density, linear settlement situated between the hills

and the Tagus River, within walking distance of the Alhandra train station. Sobralinho residents enjoy direct access to it through soft mobility, other than being well served by the Carris suburban bus network.

- Residential dormitory and reachable amenities: Sobralinho functions primarily as a bedroom community with limited local leisure amenities. Consequently, non-mandatory trips are redirected outside, generating radial flows towards the Lisbon metropolitan core. The spots that are easily accessible through the train line (Oriente and Santa Apolonia, near the city centre) are discouraged for car use due to parking constraints enforced by the EMEL. In fact, the cost for a round trip, considering fuel and parking fees, is much lower for Public Transportation. This explains why the Sobralinho outlier effect is statistically significant only for the visiting model, since leisure destinations are well served by the train line, while commuting trips target mostly industrial zones lacking direct connection with the network.

The numerical analysis to test the significance of the two aforementioned location biases will be performed in section 4.1.1.2.

After cleaning the dataset and considered spatial heterogeneity, the next step focused on selecting the most predictive variables for the final models.

3.3 Models Specification

Various statistical tools were employed to determine the right exogenous variables. The strategy involves a gradual approach, starting from bivariate analysis to test associations between mode choice and socio-demographic features (ANOVA and Cramer's V) and continuing towards dimensionality reduction techniques, such as EFA (Exploratory Factor Analysis). The analysis presented below (3.3.1) has been performed for all the considered trip purposes; however, to avoid redundancy, the variable selection process is shown in detail only for "Visiting". The methodology has been kept consistent across all remaining categories.

3.3.1 Statistical Analyses for Variable Selection

The specification of the model is an important step towards the success of the whole process, but it is very sensitive to variable selection. This translates into the identification of the right exogenous determinants within the database. The 15 minute-city focus of the survey should help to identify variables that are more categorical/ordinal than scalar. Indeed, the trip time, location and cost are not available in the dataset, thus the final utility formula must rely on Likert type or factor variables. All the variables used in this chapter are extensively explained in Appendix A1.

To identify the most significant variables to include in the final model a step-by-step process has been implemented:

- Performing Chi-Square, Cramer's V and One-way-ANOVA tests to identify associations and correlations between dependent and independent variables
- Running the model and interpreting the coefficients
- Finalize the work removing non-significant variables and avoiding non-logical patterns.

The correlations will be found through the aforementioned bivariate analyses. The tool used is "Crosstabs" of SPSS. Since most the endogenous variables are categoric, some of the tools that the software offers are not used. Therefore, for categorical variables the focalisation lies on two principal indicators, while for ordinal and scalar ones another test is employed:

3.3.1.1 Pearson Chi-Square

The Pearson Chi-Square test is used to determine whether two categorical variables are related to each other. In this case the endogenous with all the needed exogenous. The equation is the following:

$$\chi^2 = \sum_{i,j} \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}}$$

- $O_{i,j}$ are the observed frequencies in row i and column j
- $E_{i,j}$ are the expected frequencies in row i and column j

A basic requirement for the test to be significant is to have at least 5 observations in each cell of the matrix i,j.

After the χ^2 is calculated, it is compared to a critical value that depends only on the significance level required and on the degrees of freedom. For the significance level α, a standard value of 5% is adopted, while for the degrees of freedom (d.o.f.) the following formula is considered:

$$df = (i - 1)(j - 1)$$

With i the number of rows and j the number of columns.

Combining the two factors a critical value is obtained through a table and compared with the one found previously. This leads to two possible outcomes:

- If $\chi^2 < {\chi^2}_{crit}$ the null hypothesis H_0 is accepted and the two variables are unrelated.

- If $\chi^2 > \chi^2{}_{crit}$ the null hypothesis H_0 is rejected and the two variables are linked.

The χ^2 statistics is then used for the Cramer's V test.

3.3.1.2 Cramer's V test

Cramer's V is a statistical test that measures the strength of association between two variables, based on the previously calculated Chi-Square.

$$V = \sqrt{\frac{\chi^2}{N(min(i,j) - 1)}}$$

Where:

- χ^2 is the Pearson Chi-Square statistic
- N is the sample size
- i, j are respectively the number of rows and columns

Since $0 < V < 1$, the higher the value the stronger is the association. To interpret correctly the magnitude of the association, the guidelines proposed by (Cohen, 1988) were adopted. According to these benchmarks, a Cramer's V value around 0.10 is considered a small effect, 0.30 a medium effect, and 0.50 or higher a large effect.

For this reason, if $V > 0.2$, at least a moderate association is present between the two variables and should be considered for the implementation in the MNL model. The number of d.o.f. should be also considered, as the lower it is, the higher the value of Cramer's V should be in order for the correlation to be significant. Considering the outcomes presented in Table 5, a critical selection was performed taking into account the abovementioned considerations.

Table 5: Cramer's V Results for Selected Variables for Visiting Trip Purpose (Source: own elaboration using SPSS based on survey data)

Variable	p-value	Cramer's V
Age (Years)	<0.001	0.454
PT_Card (0-1)	<0.001	0.372
Driver_License_Binary (0-1)	<0.001	0.329
Education_Level_Ordinal_Opt (0-1)	<0.001	0.307
Bus_Inclination (1-5)	<0.001	0.265
Motor_Vehicles (N)	<0.001	0.207
HH_Total (N)	<0.001	0.188
Income_Range_Avg (€)	<0.001	0.145
Avg_Trip_Time_Ordinal_Opt (0-2)	0.005	0.112
Dummy_Sobr (0-1)	0.005	0.109
Density (1-5)	0.173	0.081
Parking (1-5)	0.200	0.079
Gender_Dummy (0-1)	0.068	0.078
Traffic (1-5)	0.212	0.078

3.3.1.3 One-way ANOVA

One-way ANOVA is a test that inspect if the mean of a continuous or ordinal variable differs across groups of one categorical variable. If this is the case, the variable in question is probably worth being inserted in the MNL model.

In detail, the calculation depends on mean squares between and within the groups of the categorical variable.

$$MS_{Between} = \frac{SS_{Between}}{df_{Between}}$$

$$MS_{Within} = \frac{SS_{Within}}{df_{Within}}$$

- $SS_{Between}$ = Sum of Squares between the groups
- SS_{Wit} = Sum of Squares within the groups
- k = Number of groups

- N = Number of observations
- $df_{Total} = df_{Between} + df_{Withi} \;\; = N - 1$
- $df_{Between}$ = Degrees of Freedom between the groups $= k - 1$
- df_{Within} = Degrees of Freedom within the groups $= N - k$

$$F = \frac{MS_{Between}}{MS_{Wi}}$$

The F-statistic is computed and is then compared to F_{crit}, that depends, similarly to the ${\chi^2}_{crit}$, on the d.o.f. and on the confidence level chosen, usually $\alpha = 0.05$. This leads to two possible outcomes:

- If $F < F_{crit}$ the null hypothesis H_0 is accepted, thus the means are similar in each group.
- If $F > F_{crit}$ the null hypothesis H_0 is rejected, thus the means are different across groups.

Initially observing the means through Table 6, it is evident that most of the selected variables display a very different behaviour depending on the transport mode.

Table 6: Within-Group Averages (Source: own elaboration using SPSS based on survey data)

Variable	Transport Mode	N	Mean
Income_Range_Avg (€)	Motorized	329	2398.18
	Unmotorized	280	1975.89
	Public Transport	54	2018.52
	Total	663	2188.91
Income_Range_Ordinal (0-2)	Motorized	329	1.32
	Unmotorized	280	1.06
	Public Transport	54	1.11
	Total	663	1.19
Age (Years)	Motorized	399	45.27
	Unmotorized	396	54.94
	Public Transport	88	32.83
	Total	883	48.37
Motor_Vehicles (N)	Motorized	399	1.55
	Unmotorized	396	1.20
	Public Transport	88	1.16
	Total	883	1.35
Bus_Inclination (1-5)	Motorized	399	2.99
	Unmotorized	396	3.34
	Public Transport	88	3.81
	Total	883	3.23

Residence_Years	Motorized	399	20.24
	Unmotorized	396	21.70
	Public Transport	88	19.55
	Total	883	20.82
HH_Total (N)	Motorized	399	3.02
	Unmotorized	396	2.62
	Public Transport	88	3.20
	Total	883	2.86
Avg_Trip_Time_Ordinal_Opt (0-2)	Motorized	315	1.10
	Unmotorized	214	1.11
	Public Transport	71	1.51
	Total	600	1.15
Education_Level_Ordinal_Opt (0-1)	Motorized	399	0.56
	Unmotorized	396	0.27
	Public Transport	88	0.24
	Total	883	0.40
Driver_License_Binary (0-1)	Motorized	399	0.93
	Unmotorized	396	0.70
	Public Transport	88	0.57
	Total	883	0.79
PT_Card_Binary (0-1)	Motorized	399	0.31
	Unmotorized	396	0.44
	Public Transport	88	0.95
	Total	883	0.43
Factor_Living	Motorized	399	-0.32
	Unmotorized	396	0.44
	Public Transport	88	-0.54
	Total	883	0.00
Dummy_Sobr (0-1)	Motorized	399	0.04
	Unmotorized	396	0.03
	Public Transport	88	0.10
	Total	883	0.04
Age_Ordinal (0-2)	Motorized	399	0.83
	Unmotorized	396	1.15
	Public Transport	88	0.39
	Total	883	0.93
MV_Ordinal (0-2)	Motorized	399	1.44
	Unmotorized	396	1.15
	Public Transport	88	1.11
	Total	883	1.28
Age_Ordinal_Opt (0-1)	Motorized	399	0.13
	Unmotorized	396	0.33
	Public Transport	88	0.13
	Total	883	0.22

Apart from “Residence_Years”, all the variables show noticeable shifts across the different means in a coherent way. Indeed, from the literature, it is possible to confirm that the results found are consistent with previous works. For example, it is common knowledge that motorized users tend to have a higher average income and education level, with the two variables sharply connected to each other. It is also possible to notice that older users are more inclined to walking or biking while youngsters are more prone to using public transportation.

F-statistics and related significance levels are presented in Table 7.

Table 7: ANOVA Results and Significance (Source: own elaboration using SPSS based on survey data)

Variable	F	p-value
Income_Range_Avg (€)	13.32	<0.001
Income_Range_Ordinal (0-2)	12.71	<0.001
Age (Years)	70.63	<0.001
Motor_Vehicles (N)	23.39	<0.001
Bus_Inclination (1-5)	45.30	<0.001
Residence_Years	1.00	0.37
HH_Total (N)	13.07	<0.001
Avg_Trip_Time_Ordinal_Opt (0-2)	7.13	<0.001
Education_Level_Ordinal_Opt (0-1)	45.77	<0.001
Driver_License_Binary (0-1)	53.45	<0.001
PT_Card_Binary (0-1)	70.47	<0.001
Factor_Living	85.14	<0.001
Dummy_Sobr (0-1)	5.33	0.005
Age_Ordinal (0-2)	54.54	<0.001
MV_Ordinal (0-2)	23.68	<0.001
Age_Ordinal_Opt (0-1)	29.43	<0.001

Having these results does not mean that every variable with a decent shift in the means is a significant variable for the MNL model, even if the significance is under the threshold of $\alpha = 0.05$. Of vital importance is the inspection of the actual relationship magnitude between each variable and the grouping one. The SPSS software provides a useful table called “ANOVA Effect Size”, that uses some statistical indicators.

- $\eta^2 = \frac{SS_{Between}}{SS_{Total}}$, it measures the proportion of total variance explained by group membership. It can present some biases, usually due to large values of k and small values of N, but since $k = 3$ and N is very large the actual bias is small.
- $\varepsilon^2 = \frac{SS_{Between} - df_{Between} \cdot MS_{Wi}}{SS_{Total}}$, used to reduce the bias of eta squared. In the current sample the result should be similar.
- $\omega^2 = \frac{SS_{Between} - df_{Between} \cdot MS_{Wi}}{SS_{Total} + MS_{Within}}$, it estimates the variance explained by the population and not by the sample. If the sample is adequate, it should not differ much from the two indicators above.

The results are highlighted in Table 8, as follows:

Table 8: Effect sizes of ANOVA Analysis (Source: own elaboration using SPSS based on survey data)

Variable	Eta-squared	Epsilon-squared	Omega-squared
Income_Range_Avg (€)	0.04	0.04	0.04
Income_Range_Ordinal (0-2)	0.04	0.03	0.03
Age (Years)	0.14	0.14	0.14
Motor_Vehicles (N)	0.05	0.05	0.05
Bus_Inclination (1-5)	0.09	0.09	0.09
Residence_Years	0.00	0.00	0.00
HH_Total (N)	0.03	0.03	0.03
Avg_Trip_Time_Ordinal_Opt (0-2)	0.02	0.02	0.02
Education_Level_Ordinal_Opt (0-1)	0.09	0.09	0.09
Driver_License_Binary (0-1)	0.11	0.11	0.11
PT_Card_Binary (0-1)	0.14	0.14	0.14
Factor_Living	0.16	0.16	0.16
Dummy_Sobr (0-1)	0.01	0.01	0.01
Age_Ordinal (0-2)	0.11	0.11	0.11
MV_Ordinal (0-2)	0.05	0.05	0.05
Age_Ordinal_Opt (0-1)	0.06	0.06	0.06

As it is possible to notice from the above chart, the variables that have the highest explained variance are "Age", "Bus_Inclination", "Education_Level_Ordinal_Opt", "Driver_License_Binary", "PT_Card_Binary", "Factor_Living" and "Age_Ordinal". All those variables have indicators with values near 0.1 above. From prominent guidelines established by previous studies (Cohen, 1988), it is accurate to say that the effect size is moderate to large, indicating a meaningful association between the grouping and the exogenous variables.

Concerning Motor Vehicles, the magnitude of the effect is at the limit of being considered for the final MNL model. In this case, it will be further analysed in a successive process, to understand if the variable is significant and appropriate to explain differences in transport behaviour.

3.3.2 Definition of the Latent Variable measuring the Attitudes towards 15-Minute City

In order to refine both the ICLV model and, to a lesser extent, the multinomial logit model, it is fundamental to identify the optimal latent variables (LV). These are underlying constructs that can only be inferred indirectly from measurable indicators through a mathematical model. Since most of the questions of the survey were assessed on Likert scales, it is methodologically convenient to combine them into composite factors. This process not only reduces data dimensionality, but also effectively captures the hidden preferences of the respondents regarding specific mobility themes. Furthermore, it reduces the multicollinearity issues that would arise by including all indicators directly. The first step to achieve this goal is to perform a factor analysis on the perception-based questions to build a first latent factor capturing the attitudes related to the 15-minute city.

3.3.2.1 Factor Analysis and Validation of the Extracted Dimensions

Based on the survey, four main categories of perception-based questions asked to participants were identified, which can be characterized as follows:

- Living (identified queries in the annex with code "1"): How much the person is enjoying life and participating in activities in the neighbourhood.
- Safety (identified queries in the annex with code "8"): How much the person is feeling safe when walking in the neighbourhood.
- Access (identified queries in the annex with code "2"): How much amenities are reachable within walking distance in the neighbourhood.
- Use (identified queries in the annex with code "9"): How much the person is utilizing the amenities present in the neighbourhood.

Considering the aforementioned reflection, an Exploratory Factor Analysis (EFA) was performed to determine the dominant psychological constructs and their respective weight. This analysis was executed in Python using the "factor_analyzer" package. Before the extraction, the dataset was cleaned by removing zero-variance indicators, and the Kaiser-Meyer-Olkin (KMO) test was performed to verify the sampling adequacy. It returns a value comprised between 0 and 1, which measures the proportion of variance among the variables that is common.

$$KMO = \frac{\sum_i \sum_{j\neq i} r^2{}_{i,j}}{\sum_i \sum_{j\neq i} r^2{}_{i,j} + \sum_i \sum_{j\neq i} p^2{}_{i,j}}$$

- $r_{i,j}$: Coefficient of correlation between variable i and j
- $p_{\mathrm{i,j}}$: Coefficient of partial correlation between variable i and j

The partial correlation represents the strength of relationship between two variables while holding constant the effect of another variable (the factors in the present case). As $p_{\mathrm{i,j}}$ approaches 0, it signifies that the correlation between the two variables is explained for the most part by a latent variable, increasing the KMO value. Kaiser indicates that if KMO > 0.8 the sample is adequate for dimension reduction (Kaiser & Rice, 1974). In the present case KMO $= 0.828$, thus it is possible to proceed further.

The factors were extracted using the Minimum Residuals method. This method was preferred over Maximum Likelihood (ML) in order to adequately deal with Likert-scale variables, which are not distributed normally (basis hypothesis for ML use). Furthermore, an orthogonal Varimax rotation was applied to facilitate factor interpretation by maximizing the variance between the LVs. The optimal number of latent variables was fixed to four, a decision supported by the visual interpretation of the Scree Plot illustrated in Figure 3.

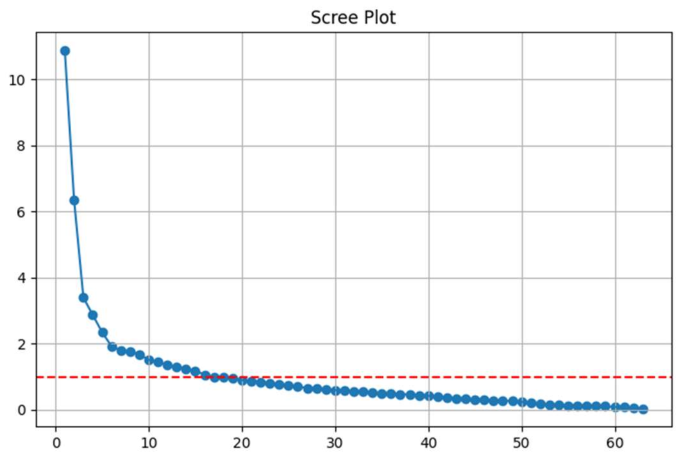

Figure 4: Scree Plot derived from EFA (Source: own elaboration)

The analysis was operated to identify 4 different global factors. The Scree Plot illustrated suggested the possibility of defining up to 6 different latent variables. In fact, according to the "elbow rule", the steepest slope change is witnessed after the sixth factor. However, after analysing preliminary weights and using common sense, it has been decided to reduce their number for two reasons:

- ➢ A smaller elbow is present around the third and fourth factors, which signifies that a second partition is also viable.
- ➢ Some of the factors found were composed by very few indicators and others were not related to each other, making it difficult to regroup them into a well-defined latent attitude.

The resulting factor structure is detailed in Figure 5 through a heatmap visualization. As usual, variables listed in the first column are defined in AppendixA1: Database Codebook.

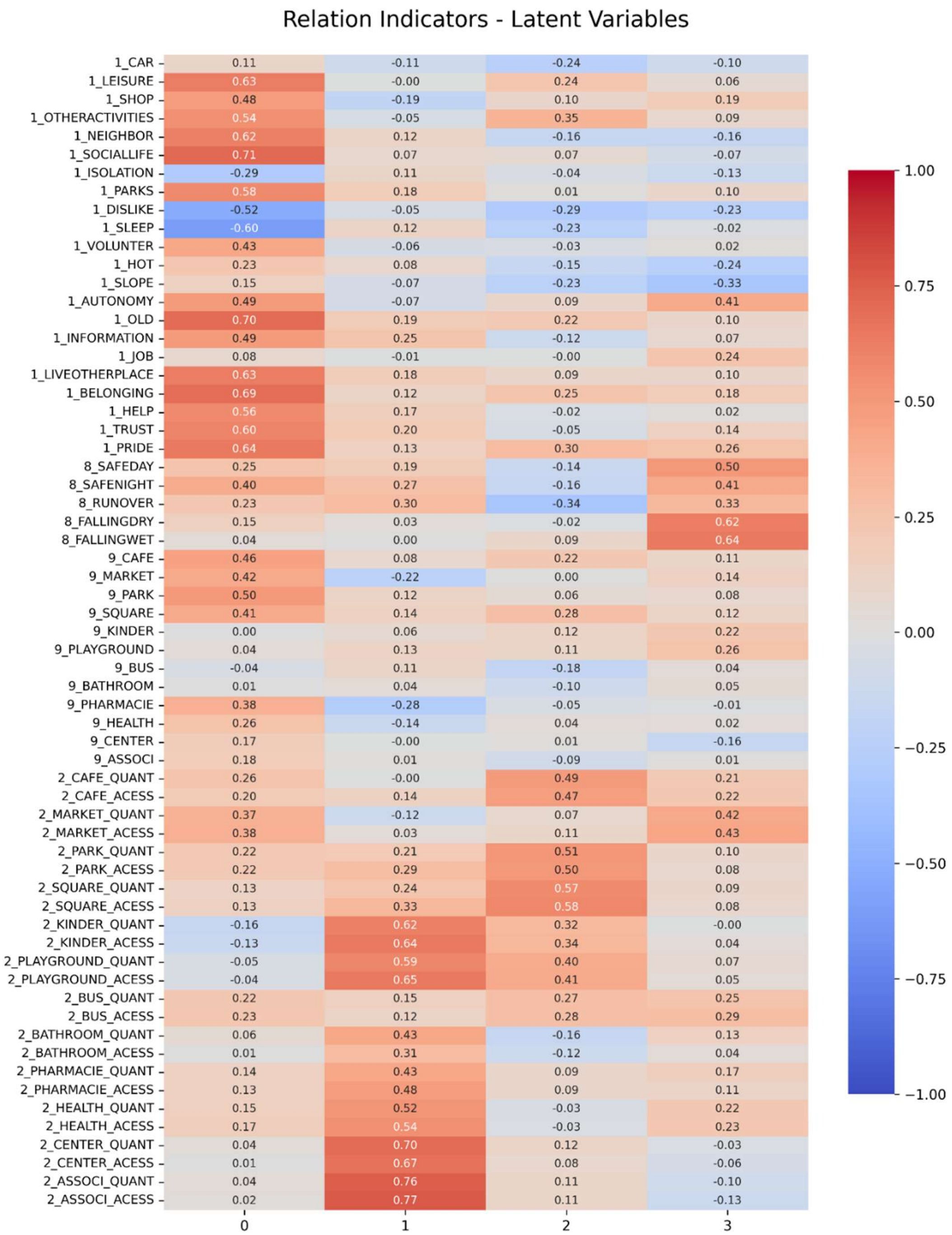

Figure 5: Exploratory Factor Analysis (Source: own elaboration)

The rows represent the variables observed in the survey, while the columns, labelled at the bottom of the table from 0 to 3, indicate the four extracted latent variables. The numerical values within each cell are the factor loadings, which quantify the correlation between the observed variables and a latent construct. The colour scale provides an intuitive representation of these loadings: dark red indicates a strong positive correlation (approaching 1), dark blue represents a strong negative correlation (approaching -1) and paler shades indicate negligible relationships (close to 0).

Knowing the heatmap attributes, it is possible to recognize clear patterns hidden in the raw data. A loading threshold of 0.40 was generally considered to establish a significant relationship between an indicator and a latent construct. That said, the dimension of neighbourhood attachment and social cohesion as factor 0 is particularly evident. It exhibits strong positive loadings on indicators related to community interaction, such as pride (variable "1_PRIDE"), neighbour trust (variable "1_TRUST") and social life (variable "1_SOCIALLIFE"). For simplicity, the name "Factor_Living_15min" was attributed to this construct. Factor 3 is identified as the perceived pedestrian safety and was named "Factor_Safety_15min". It envelops all the queries with code "9".

Concerning the two other dimensions (1 and 2), it is possible to observe that they represent a similar latent attitude: the ease of access to services and leisure (queries with code "2" in the codebook). Albeit being separated in the factor analysis into two different constructs, these variables collectively measure a single theoretical domain: the perceived accessibility and availability of local neighbourhood amenities (such as retail, healthcare and recreational facilities). Artificially splitting questions on a related topic in two categories would probably improve the mathematical accuracy of the model, but decrease the meaning of the latent attitudes. Therefore, in order to prioritize interpretability, dimensions 1 and 2 will be jointly considered as a unique "Factor_Access_15min".

The selected variables to represent each factor are shown in Table 9, maintaining the threshold specified previously.

Table 9: Factor Definition (Source: own elaboration)

Factor Name	**Variables**	
Factor_Living_15min	1_LEISURE	1_NEIGHBOR
	1_SHOP	1_SOCIALLIFE
	1_OTHERACTIVITIES	1_PARKS
	1_VOLUNTER	1_AUTONOMY
	1_OLD	1_INFORMATION
	1_LIVEOTHERPLACE	1_BELONGING
	1_HELP	1_TRUST
	1_PRIDE	
Factor_Safety_15min	8_SAFEDAY	8_FALLINGDRY
	8_SAFENIGHT	8_FALLINGWET
	8_RUNOVER	
Factor_Access_15min	2_CAFE_QUANT	2_BATHROOM_QUANT
	2_CAFE_ACESS	2_BATHROOM_ACESS
	2_PARK_QUANT	2_PHARMACIE_QUANT
	2_PARK_ACESS	2_PHARMACIE_ACESS
	2_SQUARE_QUANT	2_HEALTH_QUANT
	2_SQUARE_ACESS	2_HEALTH_ACESS
	2_KINDER_QUANT	2_CENTER_QUANT
	2_KINDER_ACESS	2_CENTER_ACESS
	2_PLAYGROUND_QUANT	2_ASSOCI_QUANT
	2_PLAYGROUND_ACESS	2_ASSOCI_ACESS

All labels shown above refer to variables that are defined in Appendix A1.

To rigorously define the latent constructs for the ICLV framework, a specific validation approach was adopted rather than a global EFA (Exploratory Factor Analysis). For each aforementioned theoretical dimension ("Factor_Living_15min", "Factor_Safety_15min", "Factor_Access_15min"), the corresponding subset of queries was isolated. A single factor extraction was then performed independently on each subset to verify its one-dimensionality, making sure that the selected items were truly explaining a single underlying psychological construct. Furthermore, the internal consistency and reliability of each construct were tested with two fit measures provided by the software: the percentage of variance explained by the factor and the value of Cronbach's Alpha. The first measures how much the total variation in a dataset is determined by the component, while the latter is calculated to measure the reliability and the internal consistency of the factorization. The results are illustrated in Table 10.

Table 10: Factors Significance (Source: own elaboration)

Factor	% of Variance Explained (EVA)	Cronbach's Alpha
Factor_Living_15min	43.789	0.878
Factor_Safety_15min	51.322	0.745
Factor_Access_15min	33.885	0.885

Generally, regarding explained variance (EVA), literature suggests, as acceptable, values of $EVA \geq 0.5$ (Shrestha, 2021). "Factor_Safety_15min" demonstrates a strong one-dimensionality, explaining 51.3% of the variance. For "Factor_Living_15min" and "Factor_Access_15min", the explained variance stands at 43.8% and 33.9%, respectively. While slightly below the 50% threshold, these values are typical when dealing with real-world human perceptions of the neighbourhood. Especially for the case of "Factor_Access_15min", the relatively lower explained variance is a direct mathematical consequence of merging the two dimensions considered by the statistical analysis. Furthermore, the high values of Cronbach's Alpha (in EFA a value of Cronbach's $\alpha > 0.7$ is considered satisfactory (Taber, 2018)), highlight the theoretical validity of the three factors.

3.3.2.2 Latent Variable Derivation

For each observation in the dataset, three scores related to the previously identified dimensions were obtained by extracting the factor scores from the manifest variables listed in Table 9. Instead of computing a simple arithmetic average, a block-wise factor extraction was performed for each construct, forcing the algorithm to extract a single factor per block of selected variables. Prior to extraction, the internal consistency of each block was validated using Cronbach's Alpha. This approach allowed the isolation of the common variance across the Likert-scale items, filtering out measurement errors and resulting in three standardized continuous indicators. Furthermore, a factor sign flipping was applied where necessary to guarantee that the three factors were leaning towards the same direction. This means that, for all three indicators, a positive score will represent a higher level of the evaluated attribute.

Since the resulting three scores are related to the general perspective of the user towards the 15-minute city, it has been decided to jointly consider them by defining one higher-order latent variable LV_1 "Attitude towards 15-min city", as evidenced in Figure 6. As shown by the ovals in this figure, the three continuous indicators have names that are derived from the corresponding factors in Table 10.

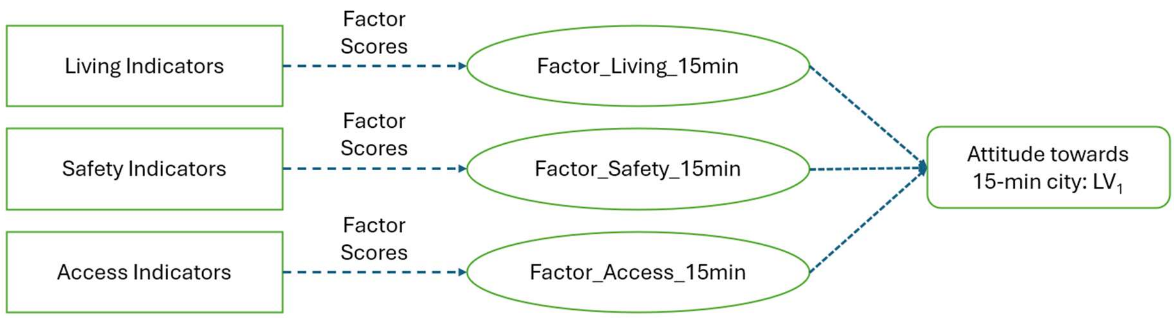

Figure 6: Determination of LV_1 through Indicators and Related Factor Scores (Source: own elaboration)

Although the standard practice suggests using the three factors as separate predictors in the model, a unique latent variable LV_1 based on scores consistent with these factors was rather introduced for two principal reasons:

- Using three latent variables was computationally heavy and would not have left space for other latent constructs.
- The results of preliminary analyses were not satisfactory. For example, the safety latent variable alone was not significant for any mode choice, while the global attitude LV_1 is a very strong predictor.

3.4 Models Framework and Software

The statistical analysis and model estimation were conducted using a dual-software approach, taking advantage of the specific strengths of different computational environments to ensure both robust data processing and flexible model specification. The preliminary analysis was performed using the SPSS (Statistical Package for the Social Sciences) software, while the most advanced was handled in the Python environment, employing the “Biogeme” library (Bierlaire, 2003).

The analysis was performed for each trip purpose introduced in section 3.1.3, maintaining the same specification in order to comment and interpret the differences between each travel behaviour.

3.4.1 The Multinomial Logit

3.4.1.1 Software Used

IBM SPSS Statistics (version 29) was employed for the preliminary stages of the analysis. Specifically, it was used for:

- Data Cleaning and Management: Handling missing values, detecting outliers and merging categories, as seen in section 3.2.2.
- Descriptive Statistics: Generating frequency tables and cross-tabulations to assess the sample distribution (3.1.3).
- Base MNL Estimation: Running initial Multinomial Logit iterations to test variable significance.

The software relies on the Newton-Raphson method to perform the iterative MLE (Maximum Likelihood Estimation) and find the optimal β coefficients. Being a root-finding algorithm, it produces successively better approximations to the solution by using the properties of the derivative. In the simplest scalar case, the formula is the following:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

Since in a Multinomial Logit model the unknown is a vector of β parameters (one for each exogenous variable), the matrix generalized form is used. Moreover, to maximize the Log-Likelihood function it is necessary to detect the point where the first derivative (gradient) is null, rather than where the function itself equals 0. The equation in matrix notation follows:

$$\beta_{n+1} = \beta_n - H^{-1}(\beta_n) \cdot\ g(\beta_n)$$

- H^{-1}: Inverse of the Hessian matrix (second derivative of Log-Likelihood)
- g: Gradient vector (first derivative of Log-Likelihood)

Specifically, the algorithm seeks the vector of β coefficients that makes the actually observed choices of the sample most probable based on the explanatory variables used in the model. By maximizing the Log-Likelihood function, the estimated parameters ensure the best possible fit to the empirical data.

Through the specific tool “Multinomial Logistic” of the “Regression” ribbon, it is possible to implement the desired model. The prompt window is displayed in Figure 7.

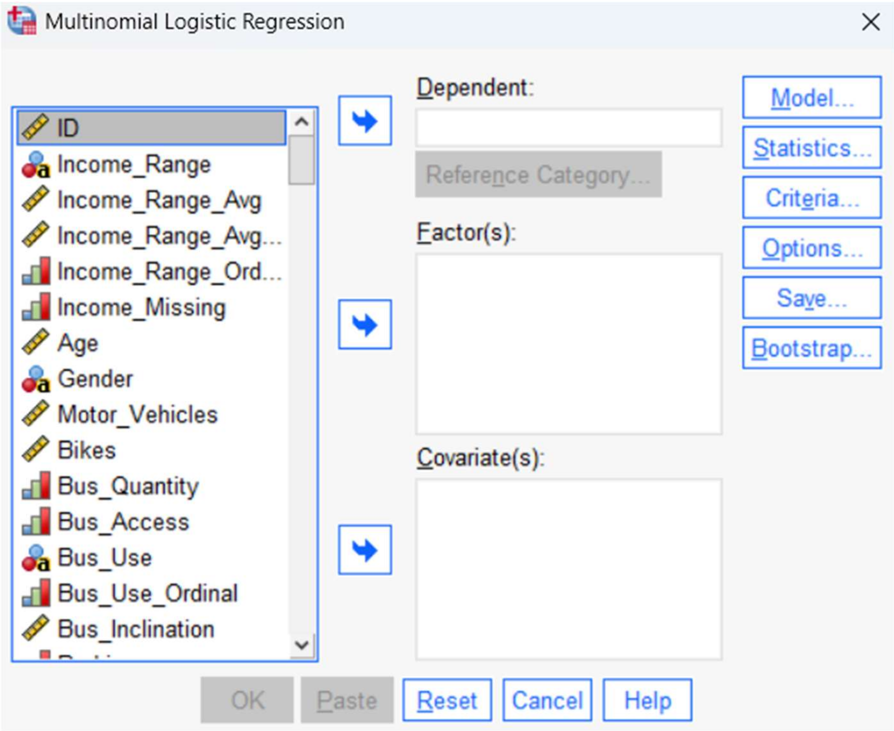

Figure 7: Prompt Window of SPSS for the Multinomial Logistic Regression (Source: own elaboration through SPSS)

The prompt window is organized into three main parts: the variable list on the left, the specification boxes in the centre and the ribbons that control the settings on the right. Concerning the main empty fields, the endogenous variable is placed in the "Dependent" box, while for the other two fields it is necessary to understand how the software distinguishes the exogenous variables:

- Factors: Treated by SPSS as qualitative variables with discrete categories, thus the model handles them by creating dummy variables. The number of the latter is $n_{dummy} = n_{categories} - 1$ since one of the classes is the reference.
- Covariates: Treated by SPSS as quantitative variables measured on a continuous scale. These are computationally more efficient, but the software fails to capture non-linear effects.

Therefore, all the Likert-scale variables have to be placed into the "Factor" field, while the metric ones have to be inserted into the "Covariate" box if the linearity assumption is deemed reasonable. The other ribbons control the mathematical execution and the printed results of the regression. In detail:

- Model: This menu allows to control model type (main effect, stepwise, etc...).

- Statistics: This controls the output tables generated by the program (pseudo R^2, classification tables, model fitting, etc…).
- Criteria: It manages the MLE (Maximum Likelihood Estimation) algorithm concerning number of iterations and converging tolerance.
- Options: This section provides tools to handle missing values or apply dispersion scaling.
- Save: It gives the possibility of saving the model results directly into the dataset as new columns.
- Bootstrap: This allows the user to perform bootstrapping (resampling the dataset thousands of times) to obtain more accurate standard errors and confidence intervals for the parameter estimates.

After the selection of the desired settings, the program estimates the regression model delivering the output viewer containing several key statistical and inferential tables. Specifically, the evaluation of the model and the interpretation of the results rely on the following outputs:

- Case Processing Summary: A preliminary table detailing the sample size (N) and the marginal percentages for each category of the dependent variable and the inserted factors. It ensures that the eventual missing data do not compromise the estimation and that the subpopulation distribution is mathematically viable, providing also a frequency table.
- Model Fitting Information and Pseudo R-Square: These tables assess the global fitness of the model. The test compares the -2 Log Likelihood of the estimated final model against a baseline "Intercept Only" model. A Chi-square test ($p < 0.05$) indicates whether the inclusion of the regressors significantly improves the model predictive power. Furthermore, Pseudo-R^2 metrics (such as McFadden, Cox, Snell and Nagelkerke) are provided as measures of the variance explained by the model, equivalent to the R^2 in linear regression.
- Likelihood Ratio Tests: This table evaluates the overall statistical significance of each independent variable across the entire model. Their individual Chi-square value is also calculated.
- Parameter Estimates: This is the core output for behavioural interpretation, detailing the specific impact of each predictor on the different transport alternatives relative to the baseline category, which is "Motorized" in the present case. For each variable, the table provides: β coefficient (log-odds), e^{β} value (odds) and statistical significance (p-value).

A positive β coefficient increases the utility of that specific alternative, holding all other variables constant. Since these parameters are expressed as log-odds, calculating their exponential gives the odds ratio. It represents the multiplicative change in the likelihood of choosing an alternative over the reference one for each one unit increase in the regressor. Finally, only predictors with a p-value $p < 0.05$ are considered statistically significant at a 95% confidence level.

3.4.1.2 Strengths and Weaknesses of the Method

The selection of the Multinomial Logit (MNL) as the preliminary analytical framework for this study is driven by three primary methodological and practical considerations:

- The specification of the model does not include time and cost of the different alternatives. Therefore, the utility of each mode depends solely on socio-demographic variables and satisfaction indicators. Information about the level of service of discarded alternatives is not required, aligning perfectly with the information available in the dataset.
- Easy to implement within SPSS for a preliminary analysis, as its closed-form probability expression allows for efficient estimation within standard statistical software packages. MNL models do not require the complex numerical integration that is necessary for example in Multinomial Probit models.
- The three main transport modes chosen for the analysis are very distinct between each other. This maximizes the likelihood that the hypothesis of IIA (Independence of Irrelevant Alternatives) is respected.

Besides the great properties and the easy mathematical manipulation, logit models come with a major drawback. The IID property discussed in section 2.2.3 leads directly to the IIA mentioned previously. Briefly, it signifies that the odds of choosing one alternative instead of another do not depend on the presence or characteristics of any other alternative. In other words, the ratio between two options' shares is constant regardless of some new opportunity entering the market. In the transportation field it is a quite unrealistic supposition, because some alternatives are more correlated than others and share unobserved attributes between them.

The ways to limit the mistakes due to this simplification are basically two:

- Relaxing the IIA implications changing the model type and adopting Nested or Mixed Logit framework.
- Checking the correlation between alternatives using purposely designed tests, like Hausman–McFadden one.

The second option is clearly preferred, principally because of the relatively high variance between the three modes chosen. IIA is usually being rejected when some of the transportation categories are very similar (Metro and Bus for example) instead.

The Hausman–McFadden test consists of identifying the differences between the coefficients of the standard model and a reduced one without one of the three categories. The same set of independent variables must be used. After extracting all the coefficients an H parameter is calculated, reporting the differences between the full and the restricted model. Here follows the detailed formula:

$$H = (\beta_r - \beta_f)^{\mathrm{T}} \, [Var(\beta_r) - Var(\beta_f)]^{-1} \, (\beta_r - \beta_f)$$

There are three different terms:

- $(\beta_r - \beta_f)^{\mathrm{T}}$: Each term is a vector that contains the coefficients of the two models, respectively the reduced and the full one. Then, the difference between the two vectors is transposed to get a row vector.
- $[Var(\beta_r) - Var(\beta_f)]^{-1}$: The two terms represent the variance-covariance matrices of the reduced and full model. Their difference is then inverted.
- $(\beta_r - \beta_f)$: Same as the first term but not transposed, in order to have a column vector instead of a row one.

Based on H value three assumptions can be made:

- $H > 0;\ H > \chi^2{}_{0.95}$: Reject IIA
- $H > 0;\ H < \chi^2{}_{0.95}$: Cannot reject IIA
- $H < 0$: Cannot reject IIA

The results of Hausman-McFadden test will be investigated in section 4.1.3.

3.4.1.3 Model Specification

Another constraint of the SPSS software stands in the limitations towards utility functions specification. Indeed, in standard SPSS outputs, the same vector of exogenous variables is used to predict the probability of choosing a certain mode relative to the reference category. The problem stands in the impossibility of inserting specific variables for certain alternatives only, which standardizes all the utility functions. Differently, the Python environment allows a complete customization of the utility functions, making them unique and alternative specific.

Regarding model specification, the chosen reference category of the endogenous variables was "Motorized", thus all the coefficients (log-odds relative to the baseline) presented in chapter 4 are relative to the abovementioned mode. The predictors with the highest degree of association identified with Cramer's V and ANOVA tests (3.3.1) were initially included into the Multinomial Logit as covariates and factors. Other variables and various recoding variants were also tested for completeness, as it is possible to appreciate in Table 11, where variables tested but finally excluded from the model specification are listed in the right column. The reader is referred to appendix A1 for a definition of all these variables.

Table 11: Variables Specification for MNL Model (Source: own elaboration)

Included Variables ($p<0.05$)	**Excluded Variables ($p>0.05$)**
Factor_Living (Covariate)	Income_Range_Avg (Covariate)
MV_Ordinal_Opt (Factor)	Age (Covariate)
Age_Ordinal (Factor)	Factor_Safety (Covariate)
Driver_License_Binary (Factor)	Factor_Access (Covariate)
PT_Card_Binary (Factor)	Motor_Vehicles (Covariate)
Education_Level_Ordinal_Opt (Factor)	Bus_Inclination (Covariate)
Clean_Opt (Factor)	Income_Range_Ordinal (Factor)
Traffic_Opt (Factor)	Gender_Dummy (Factor)
Dummy_Sobr (Factor)	Residence_Type_Binary (Factor)
Dummy_Povoa (Factor)	Avg_Trip_Time_Ordinal_Opt (Factor)

The final MNL specification is therefore implying the following systematic utility of mode i:

$$
\begin{aligned}
V_i = ASC_i &+ \beta_{Factor\_Living,i} \cdot Factor\_Living + \beta_{MV\_Ordinal\_Opt,i} \cdot MV\_Ordinal\_Opt \\
&+\beta_{Age,i} \cdot Age + \beta_{PT\_Card\_Binary,i} \cdot PT\_Card\_Binary \\
&+\beta_{Driver\_License\_Binary,i} \cdot Driver\_License\_Binary \\
&+\beta_{Educ\_Level\_Ordinal\_Opt,i} \cdot Education\_Level\_Ordinal\_Opt \\
&+\beta_{Clean\_Opt,i} \cdot Clean\_Opt + \beta_{Traffic\_Opt,i} \cdot Traffic\_Opt \\
&+\beta_{Dummy\_Sobr,i} \cdot Dummy\_Sobr + \beta_{Dummy\_Povoa,i} \cdot Dummy\_Povoa
\end{aligned}
$$

Where ASC_i is the Alternative Specific Constant for mode i. The other variables are defined in appendix A1 and appear in the above linear formulation of V_i along with their β coefficients, which have to be estimated.

3.4.2 Hybrid Choice Model and the Python Libraries

3.4.2.1 Software Used

The Python programming environment was utilized for the advanced modelling phase (A2: Python Code). Due to the computational complexity of the Integrated Choice and Latent Variable

(ICLV) model, which requires the simultaneous estimation of utility functions and structural equations, SPSS is in fact no longer sufficient and more advanced tools are needed.

The Python's software package "Biogeme 3.3.1" has been designed for the maximum likelihood estimation of discrete choice models. It is a fully integrated and open-source Python library, and it is considered one of the academic standards for transportation modelling. Combining the aforementioned tool with the "Pandas" library for data management permits the definition of custom utility functions (unlike in SPSS) and the integration of the latent factors.

3.4.2.2 Mathematical Characterization

One of the peculiarities of the ICLV model is that the latent variable model and the mode choice model must be estimated simultaneously. Since the inclusion of unobserved latent variables leads to calculations of integrals without a closed-form solution, the estimation relies on the Maximum Simulated Likelihood (MSL) method. The integrals are managed through the Monte Carlo simulation, which performs multiple iterations to reach convergence. It approximates an integral by taking the average of random tries, called "draws". For each of them, the program assigns a value for the latent variable, that influence both indicators and modal choice. It, then, evaluates if the result obtained is consistent with the actual choice of the user, giving a probability: $P(Choice|LV)$. The same is done with indicators. All the draws are subsequently averaged, and the software tries to maximize the following probability:

$$P(Combined) = P(Choice|LV) \cdot P(Indicators|LV)$$

All the parameters are adjusted iteratively until the algorithm finds the values that are maximizing the probability across the whole population.

For the simulation process, Halton draws were employed. Compared to random ones, they provide a more uniform coverage of the integration domain compared to standard Monte Carlo methods. The latter, being completely aleatory, can simulate stack of values that are very close to each other, while Halton sequences fill the space evenly, requiring less draws to achieve the same precision.

That said, Halton draws converge significantly faster than independent ones. Literature (Bhat, 2001; Sándor & Train, 2004) affirms that 100 Halton draws can provide a level of accuracy comparable to 1000 random draws. This is critical for an ICLV model, which is computationally heavy to estimate. Reducing by an order of magnitude the number of simulations that the program has to perform leaves room for more complex operations, which consist in a larger amount of variables or the estimation through Monte-Carlo simulations for the structural part.

Indeed, for the measurement equations of the indicators (to be presented in subsection 3.4.2.4), a continuous linear specification with Gaussian errors was adopted instead of an ordered probability model (such as ordered Probit or Logit). This methodological approach is justified by the specific nature of the input data used for the different constructs. Specifically, for the first latent variable (LV_1), the inputs are continuous factor scores. Conversely, for the other construct (LV_2), the indicators are based on 5-point Likert scales. Although Likert items are ordinal, treating 5-point scales as continuous variables is an accepted practice in structural equation modelling and will be justified in the next section (3.4.2.3). This approximation significantly reduces the computational weight of the model while preserving reliable results. Furthermore, the present choice provides consistency with the structural definition of the latent variables. Since the latent constructs are assumed to follow a standard Normal distribution ($LV \sim N(0,1)$), modelling the measurement error terms with a Gaussian distribution keeps consistency with these properties. Therefore, while the Choice Model employs the Multinomial Logit formulation, thus assuming Gumbel-distributed errors (see section 2.2.1), the structural part has been treated under the assumption of normality for mathematical coherence.

3.4.2.3 Latent Variables Determination and Related Structural Relationships

Prior to the specification of the Integrated Choice and Latent Variable (ICLV) model, the latent constructs had to be defined. In fact, the process of exogenous variables choice for the utility equation is analogous to the MNL model. One of the latent constructs has already been identified (3.3.2.2). However, to accurately represent the user interaction with the 15-minute city framework, it was chosen to specify three distinct Latent Variables (LV), since a single latent construct would have been insufficient. Indeed, considering the dynamics explored in section 2.4, it was assumed that travel behaviour was influenced by at least three different dimensions: the personal attitude towards the living area, the external physical perception of the road system and the quality of pedestrian infrastructure. The latter was lately excluded from the final specification due to significance issues to preserve model parsimony, but its role will be analysed nonetheless (4.2.3.2). Therefore, combining only the crucial latent variables into the final utility formula will provide a better understanding about people hidden behaviours.

As discussed in section **Errore. L'origine riferimento non è stata trovata.**, the measurement equations (indicators) for LV_1 representing the attitude of the user towards the 15-minute city were directly derived from the Exploratory Factor Analysis (EFA), which successfully grouped the psychological constructs based on the answers to Likert-type questions.

Regarding the regressors related to LV_1, the formulation relies on one primary socio-demographic trait: age. As opposed to the previous interpretation about the non-linear effect of this parameter on mode choice, for the definition of the latent variable it was decided to keep a linear relationship. Indeed, there is no reason to think that the latter exhibits a U-shaped correlation, principally because, as individuals age, their capability to move naturally diminishes. Therefore, the reduction in physical mobility reinforces the need for close amenities, perfectly aligning with the 15-minute city framework. Unlike its effect on car use, it is reasonable to assume that the relationship between age and this attitude is linear. At last, the equations of the LVs should be kept as light as possible to ensure model convergence, principally because of their high computational weight.

Therefore, the structural relationship to compute LV_1 is the following:

$$LV_1 = \beta_{0,1} + \beta_{Age,1} \cdot Age + \varepsilon_1$$

Where $\beta_{0,1}$ is the intercept of LV_1, representing the baseline of the latent construct.

An additional latent variable, LV_2 or "Perceived Easiness of Car Use", was introduced to consider the user perception regarding the attractiveness of the motorized environment in the neighbourhood. This construct evaluates how the private transport is easy to exploit within the urban space. The related measurement indicators were directly selected from the following two variables available in the survey dataset ,thus without the need to reduce the dimensionality of the data through factor analysis as done for LV_1 in section **Errore. L'origine riferimento non è stata trovata.**:

- Satisfaction towards intensity of traffic: "Traffic"
- Satisfaction towards the availability and ease of parking: "Parking"

Methodologically, although the LV_1 indicators were treated as continuous factor scores, the LV_2 related ones ("Traffic" and "Parking") are raw 5-point Likert scales (as shown in A1: Database Codebook). However, they were assumed to be quasi-continuous and modelled with a gaussian specification. This procedure relies on the assumption that the perception of physical quantities, such as vehicular flow or parking availability, can be approximated to a continuous distribution. Moreover, in the methodological literature it is possible to find evidence that treating an ordinal Likert-scale variable with 5 or more categories as continuous gives robust results (Rhemtulla et al., 2012). This allows also a reduction in model complexity and improves computational efficiency without compromising the structural coefficients.

Concerning the structural equation for LV_2, the final specification does not include any socio-demographic regressors. This is the outcome of an iterative preliminary testing phase during which several attributes (such as age, gender, and education) were evaluated. Since none of them reached an acceptable level of statistical significance, LV_2 was defined purely by its baseline intercept and the normally distributed error term. This implies that the perception and sensitivity towards traffic congestion and parking availability are assumed not to depend on specific demographic segments, but are a psychological trait that is evenly distributed across the population.

Consequently:

$$LV_2 = \beta_{0,2} + \varepsilon_2$$

Where $\beta_{0,2}$ is the intercept of LV_2, representing the baseline of the latent construct.

Originally, the framework included a third latent variable (LV_3), which was aiming to capture the dimension of the pedestrian infrastructure. It was labelled "Quality of Pedestrian Environment" and the selected indicators for the measurement equation were four:

- Tree Availability
- Bench Availability
- Lighting Availability
- Sidewalk Quality.

However, this variable was omitted from the final model specification as the resulting coefficients were consistently low and lacked statistical significance across all evaluated trip purposes. Apart from being a technical necessity to ensure model parsimony, the exclusion of LV_3 offers diverse interesting behavioural insights. Comparing it with the high significance of LV_1, it suggests a presence of a hierarchy of factors leading to the adoption of active mobility (Alfonzo, 2005). While the first latent construct encompasses a global attitude towards the idea of 15-minute city, the third focuses on the physical structures of the road and its aesthetic improvements. Therefore, excluding LV_3 from the unmotorized utility reinforces the role of LV_1 as the primary driver towards active mobility in this sample, suggesting that residents prioritize the quality of residential environment over urban design. This aspect will be later discussed in section 4.2.3.2.

3.4.2.4 Model Specification

Concerning the specification, the systematic utilities of the three alternatives remain consistent with the baseline Multinomial Logit (MNL) model. The only structural modification is the inclusion of the latent variables as additional explanatory variables. This comparative approach is specifically adopted to evaluate and justify the added explanatory power provided by the psychometric constructs. Based on the aforementioned considerations, the structure of the ICLV model is illustrated in Figure 8.

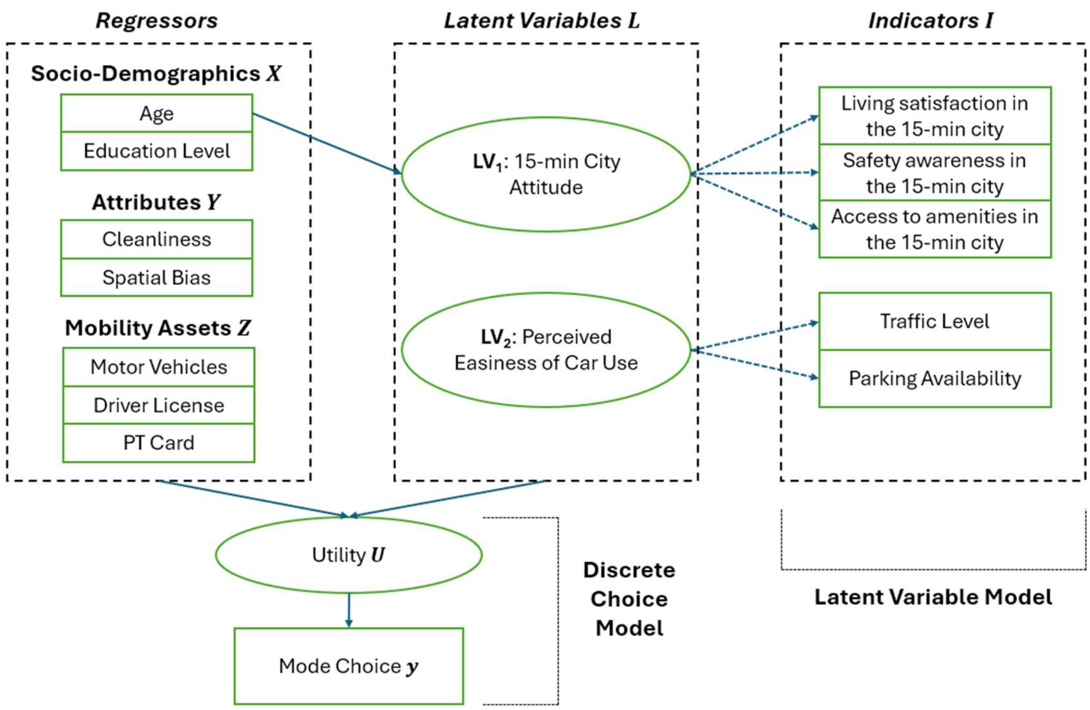

Figure 8: Path Diagram of the ICLV Model (Source: own elaboration)

As depicted in the graph, in a hybrid model both the regressors and the latent variables can be chosen as factors in the utility functions. Obviously, having three different modes, the systematic utility formula should have different parameters accounting for each mode specificity. For example, the possession of the public transport pass is an essential variable for V_{PT} but is not significative for the other two modes.

Based on the aforementioned considerations, the ICLV specification can be formalized as follows:

$$\begin{aligned} V_{Mot} = {} & \beta_{Education\_Level\_Ordinal\_Opt,Mot} \cdot Education\_Level\_Ordinal\_Opt \\ & +\beta_{MV\_Ordinal,Mot} \cdot MV\_Ordinal \\ & +\beta_{Driver\_License\_Binary,Mot} \cdot Driver\_License\_Binary \end{aligned}$$

$$\begin{aligned} V_{Unm} = {} & ASC_{Unm} + \beta_{LV_1,Unm} \cdot LV_1 + \beta_{LV_2,Unm} \cdot LV_2 \\ & +\beta_{Age,Unm} \cdot Age + \beta_{Age^2,Unm} \cdot Age^2 \\ & +\beta_{Dummy\_Povoa,Unm} \cdot Dummy\_Povoa \end{aligned}$$

$$V_{PT} = ASC_{PT} + \beta_{Age,PT} \cdot Age + \beta_{Age^2,PT} \cdot Age^2$$

$$+\beta_{PT_{Card\_Binary},PT} \cdot PT\_Card\_Binary$$
$$+\beta_{Clean\_Opt,\,PT} \cdot Clean\_Opt$$
$$+\beta_{Dummy\_Sobr,PT} \cdot Dummy\_Sobr$$

Where ASC_{Unm} and ASC_{PT} are respectively the Alternative Specific Constants of "Unmotorized" and "Public Transport" modes.

3.4.2.5 Model Identification Strategy

The estimation of Integrated Choice and Latent Variable (ICLV) models requires the imposition of specific constraints to solve its intrinsic indeterminacy. In fact, in the measurement part, the program has to estimate the factor loading and the latent variable value at the same time. LV's are theoretical and abstract concepts, which don't have a defined measurement unit. Therefore, an identification problem exists because infinite combinations of λ_k and $LV_{l,n}$ can give the same value for indicator $I_{k,n,l}$.

The possible solutions are two:

- Normalize to 1 the factor loading λ_k associated with the first indicator for each latent variable.
- Set the error variance to 1 and the mean to 0: $\varepsilon_l \sim N(0,1)$. This normalization fixes the scale of the latent variable.

Being the academic standard in ICLV modelling, the second option was chosen.

At this point, all the preliminary steps are completed. Therefore, the model can be estimated and its results commented.

4. Results

4.1 Base Model Results: Multinomial Logit

4.1.1 Preliminary Data Validation

Before investigating the results of the Multinomial Logit model, it is crucial to validate the dataset to prevent estimation biases. Specifically, two preliminary analyses were conducted: the Mann-Whitney test to assess the missing data mechanism for the income variable and a location bias test to address spatial heterogeneity across the survey locations.

4.1.1.1 Mann-Whitney Test Results

In order to investigate the existence of eventual systematic patterns behind the missing income data emerged in section 3.2.2, six exogenous variables were analysed to detect the magnitude of the related non-response bias. Extensively explained in A1, these covariates have yielded the results highlighted in Table 12.

Table 12: Mann-Whitney Test Result on Selected Variables (Source: own elaboration through SPSS)

Variable	Income_Missing	N	Mean Rank	p-value
Age	0	279	525.16	0.86
	1	776	529.02	
Education_Level_Ordinal_Opt	0	279	482.04	<0.001
	1	776	544.53	
Motor_Vehicles	0	279	520.87	0.62
	1	776	530.56	
PT_Card_Binary	0	279	582.66	<0.001
	1	776	508.35	
Driver_License_Binary	0	279	512.11	0.14
	1	776	533.71	
Gender_Dummy	0	279	504.57	0.08
	1	776	536.42	

The results indicate that the missing values of income are not entirely randomly distributed across the sample. While no significant differences were found regarding Age ($p = 0.86$), Motor Vehicle ownership ($p = 0.62$) and Driver License ($p = 0.14$) significant disparities emerged in education and for owners of a public transport pass. Moreover, gender exhibit a marginal significance that cannot be totally overlooked. These findings suggest a possible MNAR mechanism, where the likelihood of answering the income question is related to the user socioeconomic profile and mobility habits.

Therefore, to avoid introducing a bias in the final model specification, the income variable was excluded from the list of candidate exogenous variables in our models. Instead, education level and vehicle ownership can be used as robust proxies for the user SES (Socio-Economic Status), as they have almost no missing observations and high explanatory power for the financial situation of the respondent.

4.1.1.2 Location Bias Test Results

To verify if the specific interview locations mentioned in 3.2.4 induced an unobserved spatial bias in mode choice, the survey locations were temporarily inserted as explanatory variables in a preliminary MNL estimation. "Alverca do Ribatejo" was maintained as the spatial reference category. As it is possible to see in Table 13, both the coefficients of "Sobralinho" and "Póvoa de Santa Iria" are significant for the respective modes. The two have a significance $p < 0.05$, so statistically significant at the 95% confidence level. The other coefficients are, on the other hand, producing only statistical noise, as is seen below.

Table 13: Coefficients and Significance of the Location Dummies (Source: own elaboration through SPSS)

Location	**Public Transport**		**Unmotorized**	
	B	**p-value**	**B**	**p-value**
Castanheira do Ribatejo	0.253	0.647	0.645	0.110
Vila Franca de Xira	-0.186	0.673	-0.065	0.827
Alhandra	-0.651	0.305	0.393	0.246
Vialonga	0.483	0.361	0.187	0.578
Póvoa de Santa Iria	-0.381	0.450	1.616	<0.001
Forte da Casa	0.065	0.919	0.334	0.318
Sobralinho	1.639	0.013	0.195	0.690
Alverca do Ribatejo	Ref.	-	Ref.	-

The best strategy at this point is the removal of all the dummies but the ones that are related to the two above mentioned neighbourhoods. Concretely, two binary variables "Dummy_Sobr" and "Dummy_Povoa" were created with the following specification: if the survey questions were gathered in those locations, a value of "1" was assigned, otherwise 0. Subsequently, the new variables were inserted as exogenous into the final model and both maintained their significance, thus explaining the singular spatial variability of the two specific locations: Povoa de Santa Iria towards unmotorized mobility and Sobralinho towards public transport.

4.1.2 MNL Estimation Results

To better understand behavioural differences across trip purposes, four distinct Multinomial Logit models were computed: Commuting, Visiting, Shopping and 15-minute. For all estimations, the "Motorized" alternative was maintained as the reference category and the variables used were identical. This ensures consistency in the interpretation of the log-odds within different trip purposes.

4.1.2.1 Overall Model Performance: Goodness of Fit

Before examining the individual parameter estimates, it is essential to evaluate the global predictive power of each model. Table 14 summarizes the goodness-of-fit metrics across the four scenarios.

Table 14: Goodness of Fit Parameters across Trip Purposes (Source: own elaboration through SPSS)

Metric	**Commuting**	**Visiting**	**Shopping**	**15-min**
Valid Observations (N)	700	883	995	1053
Number of Estimated Parameters (k)	22	22	22	22
Model Chi-Square (χ^2)	527.38	523.86	439.83	237.01
Null Log-Likelihood (LL(0))	-580.13	-837.43	-878.10	-805.03
Final Log-Likelihood (LL(β))	-316.44	-575.51	-658.19	-686.53
McFadden Pseudo R^2	0.455	0.313	0.250	0.147
McFadden Pseudo R^2 Adjusted	0.417	0.287	0.225	0.120

Even if all models are globally significant, the McFadden Pseudo R^2 reveals a clear hierarchy in the explanatory power of the socio-demographic and environmental variables. The commuting model exhibits an excellent fit (highest Chi-Square and R^2), confirming that mandatory trips are highly structured and the most predictable. These are based principally on objective constraints, such as vehicle ownership. Conversely, the discretionary trips (visiting and shopping) show progressively lower pseudo-R^2 values, while the 15-minute area model has the lowest fit (lowest Chi-Square and R^2). This decreasing trend supports the initial hypothesis: non-mandatory and localized trips are less constrained by rigid socio-demographic factors and are likely influenced by unobserved psychological constructs, which will be later considered in the ICLV framework.

4.1.2.2 Parameter Estimates

By cross analysing the parameter estimates across the four trip purposes, a general overview of travel behaviour emerges. Instead of evaluating each model in isolation, comparing the shifts in statistical significance and coefficient magnitudes among the different models reveals several interesting dynamics. The results are, thus, presented sequentially in Table 15, Table 16, Table 17 and Table 18. Exogenous variables categories shown in the different tables are defined in A1.

Table 15: Parameter Estimates of Commuting Trip Purpose (Source: own elaboration through SPSS)

Variable	**β**	**Std. Error**	**p-value**
Public Transport (Reference: Motorized)			
Intercept	0.348	1.181	0.769
Factor_Living	-0.080	0.144	0.577
Education_Level_Ordinal_Opt=0 (Ref=1)	0.647	0.288	0.025
Driver_License_Binary=0 (Ref=1)	1.937	0.447	<0.001
PT_Card_Binary=0 (Ref=1)	-4.116	0.463	<0.001
Age_Ordinal=0 (Ref=2)	2.230	0.942	0.018
Age_Ordinal=1 (Ref=2)	1.154	0.928	0.214
MV_Ordinal_Opt=0 (Ref=1)	2.573	0.715	<0.001
Dummy_Sobr=0 (Ref=1)	-0.476	0.686	0.487
Clean_Opt=0 (Ref=1)	-0.657	0.345	0.057
Traffic_Opt=0 (Ref=1)	-0.984	0.301	0.001
Dummy_Povoa=0 (Ref=1)	-0.989	0.346	0.004
Unmotorized (Reference: Motorized)			
Intercept	-4.262	1.497	0.004
Factor_Living	0.533	0.182	0.003
Education_Level_Ordinal_Opt=0 (Ref=1)	1.156	0.352	0.001
Driver_License_Binary=0 (Ref=1)	1.768	0.470	<0.001
PT_Card_Binary=0 (Ref=1)	-0.137	0.390	0.726
Age_Ordinal=0 (Ref=2)	-0.062	0.807	0.939

Age_Ordinal=1 (Ref=2)	-0.408	0.745	0.584
MV_Ordinal_Opt=0 (Ref=1)	2.935	0.780	<0.001
Dummy_Sobr=0 (Ref=1)	0.393	1.090	0.718
Clean_Opt=0 (Ref=1)	1.147	0.557	0.039
Traffic_Opt=0 (Ref=1)	0.352	0.358	0.326
Dummy_Povoa=0 (Ref=1)	-0.235	0.368	0.524

Table 16: Parameter Estimates of Visiting Trip Purpose (Source: own elaboration through SPSS)

Variable	**β**	**Std. Error**	**p-value**
Public Transport (Reference: Motorized)			
Intercept	0.157	0.914	0.864
Factor_Living	-0.241	0.157	0.125
Education_Level_Ordinal_Opt=0 (Ref=1)	0.640	0.352	0.069
Driver_License_Binary=0 (Ref=1)	1.291	0.357	<0.001
PT_Card_Binary=0 (Ref=1)	-2.890	0.552	<0.001
Age_Ordinal=0 (Ref=2)	0.963	0.460	0.037
Age_Ordinal=1 (Ref=2)	-0.672	0.481	0.162
MV_Ordinal_Opt=0 (Ref=1)	1.922	0.511	<0.001
Dummy_Sobr=0 (Ref=1)	-1.461	0.626	0.020
Clean_Opt=0 (Ref=1)	-0.688	0.335	0.040
Traffic_Opt=0 (Ref=1)	-0.741	0.321	0.021
Dummy_Povoa=0 (Ref=1)	0.361	0.450	0.422
Unmotorized (Reference: Motorized)			
Intercept	1.005	0.613	0.101
Factor_Living	0.853	0.109	<0.001
Education_Level_Ordinal_Opt=0 (Ref=1)	0.560	0.190	0.003
Driver_License_Binary=0 (Ref=1)	1.077	0.271	<0.001
PT_Card_Binary=0 (Ref=1)	-0.235	0.198	0.236
Age_Ordinal=0 (Ref=2)	-0.415	0.309	0.180
Age_Ordinal=1 (Ref=2)	-0.351	0.236	0.138
MV_Ordinal_Opt=0 (Ref=1)	2.203	0.443	<0.001
Dummy_Sobr=0 (Ref=1)	0.017	0.468	0.971
Clean_Opt=0 (Ref=1)	-0.011	0.212	0.959
Traffic_Opt=0 (Ref=1)	-0.366	0.183	0.046
Dummy_Povoa=0 (Ref=1)	-1.493	0.204	<0.001

Table 17: Parameter Estimates of Shopping Trip Purpose (Source: own elaboration through SPSS)

Variable	**β**	**Std. Error**	**p-value**
Public Transport (Reference: Motorized)			
Intercept	-2.885	1.102	0.009
Factor_Living	-0.010	0.160	0.948
Education_Level_Ordinal_Opt=0 (Ref=1)	0.563	0.445	0.206
Driver_License_Binary=0 (Ref=1)	1.305	0.362	<0.001
PT_Card_Binary=0 (Ref=1)	-2.820	0.473	<0.001
Age_Ordinal=0 (Ref=2)	0.359	0.436	0.411
Age_Ordinal=1 (Ref=2)	-0.575	0.398	0.149
MV_Ordinal_Opt=0 (Ref=1)	3.085	0.459	<0.001
Dummy_Sobr=0 (Ref=1)	0.363	0.699	0.604
Clean_Opt=0 (Ref=1)	0.025	0.352	0.943
Traffic_Opt=0 (Ref=1)	-0.638	0.332	0.055
Dummy_Povoa=0 (Ref=1)	0.787	0.652	0.227
Unmotorized (Reference: Motorized)			
Intercept	-0.201	0.625	0.748
Factor_Living	0.353	0.091	<0.001
Education_Level_Ordinal_Opt=0 (Ref=1)	0.195	0.176	0.267
Driver_License_Binary=0 (Ref=1)	0.761	0.236	0.001
PT_Card_Binary=0 (Ref=1)	-1.376	0.173	<0.001
Age_Ordinal=0 (Ref=2)	0.100	0.274	0.715
Age_Ordinal=1 (Ref=2)	-0.352	0.210	0.094
MV_Ordinal_Opt=0 (Ref=1)	2.025	0.389	<0.001
Dummy_Sobr=0 (Ref=1)	1.454	0.524	0.005
Clean_Opt=0 (Ref=1)	0.014	0.193	0.944
Traffic_Opt=0 (Ref=1)	0.127	0.168	0.448
Dummy_Povoa=0 (Ref=1)	-1.333	0.192	<0.001

Table 18: Parameter Estimates of 15-Minute Trip Purpose (Source: own elaboration through SPSS)

Variable	**β**	**Std. Error**	**p-value**
Public Transport (Reference: Motorized)			
Intercept	-18.305	0.862	<0.001
Factor_Living	0.090	0.160	0.575
Education_Level_Ordinal_Opt=0 (Ref=1)	0.788	0.435	0.070
Driver_License_Binary=0 (Ref=1)	-0.039	0.399	0.923
PT_Card_Binary=0 (Ref=1)	-3.550	0.746	<0.001
Age_Ordinal=0 (Ref=2)	0.742	0.450	0.099

Age_Ordinal=1 (Ref=2)	-0.582	0.429	0.175
MV_Ordinal_Opt=0 (Ref=1)	1.984	0.495	<0.001
Dummy_Sobr=0 (Ref=1)	16.935	0.000	<0.001
Clean_Opt=0 (Ref=1)	-0.361	0.362	0.318
Traffic_Opt=0 (Ref=1)	-0.513	0.335	0.126
Dummy_Povoa=0 (Ref=1)	0.776	0.642	0.227
Unmotorized (Reference: Motorized)			
Intercept	1.519	0.500	0.002
Factor_Living	0.497	0.083	<0.001
Education_Level_Ordinal_Opt=0 (Ref=1)	0.243	0.169	0.150
Driver_License_Binary=0 (Ref=1)	0.199	0.249	0.424
PT_Card_Binary=0 (Ref=1)	-0.433	0.177	0.014
Age_Ordinal=0 (Ref=2)	0.883	0.288	0.002
Age_Ordinal=1 (Ref=2)	0.161	0.214	0.451
MV_Ordinal_Opt=0 (Ref=1)	1.409	0.402	<0.001
Dummy_Sobr=0 (Ref=1)	0.146	0.351	0.676
Clean_Opt=0 (Ref=1)	-0.125	0.195	0.523
Traffic_Opt=0 (Ref=1)	-0.348	0.163	0.033
Dummy_Povoa=0 (Ref=1)	-0.701	0.210	0.001

The most evident patterns emerging from the coefficients of the variables and their significance across the four considered models can be grouped into six distinct macro-categories:

- The living factor impact on unmotorized mobility: it is noticeable that the influence of "Factor_Living" is consistent for active travel across all trip purposes. Albeit a tangible change in magnitude from visiting (β=0.853) to shopping (β=0.353), the coefficient remains positive and significant regardless of the travel function. Users who experience a higher neighbourhood satisfaction tend to favour active travel modes. This provides clear quantitative evidence that good urban design universally promotes active mobility, independently of the trip purpose.
- The mobility assets importance: not owning a public transport pass or a driver license affects massively the probability of choosing respectively transit or motorized mobility. For the variable "PT_Card_Binary", even if the sign is constantly negative, the intensity ranges from β=-4.116 for commuting to β=-2.820 for shopping, principally due to the lack of flexibility of mandatory trips. Similarly, having a driver license influences positively the utility of motorized mode, apart from the 15-minute purpose. Indeed, for short trips, users with a license behave similarly to those without one. This indicates the efficiency of active modes for the 15-minute city framework. Conversely, owning a vehicle acts as a strong

factor towards motorized mode in every model. This confirms the role of “captive users” in choosing alternative modes to the car, lacking a real choice (2.4.1). Indeed, the possession of mobility assets is a substantial determinant in mode choice.

- Socio-demographic variables affecting differently mandatory and discretionary trips: education level plays an important role, but only for certain trip purposes. Indeed, for commuting, individuals without a university degree are significantly more likely to use public transport (β=0.647) and active modes (β=1.156) instead of driving. This does not necessarily imply that educated people are less concerned about sustainability concerns, but could be directly related to the socio-economic level (2.4.1), with the previously mentioned variable acting as a proxy for financial capability. Regarding age, the general pattern reveals a higher propensity for public transport by the young users (<30), especially for commuting (β=2.230) and marginally for visiting (β=0.963). Interestingly, in the 15-minute framework, data indicates a strong reliance on active modes among the younger demographic class, probably because of better walking ability and overall health (2.4.1).
- Spatial attributes heterogeneity across trip purposes: the inclusion of location dummies highlights how local geography impacts specific trip purposes. For example, living outside Póvoa de Santa Iria has a strongly negative and significant effect on walking for visiting (β=-1.493), shopping (β=-1.333), and 15-minute trips (β=-0.701). This suggests that the location possesses an urban layout that is highly favourable to active mobility for discretionary purposes, as previously mentioned in section 3.2.4. However, this negative coefficient reduces significantly for mandatory commuting trips (β=-0.235, p-value>0.05). It could imply that, while local terrain or infrastructure in certain neighbourhoods could encourage active mobility, commuters will use a private vehicle regardless of these spatial advantages if required by their daily routine. Regarding the other spatial dummy (Dummy_Sobr), a similar observation to the one seen in section 3.2.4 can be proposed. This implies a high local reliance on public transport, especially for discretionary purposes, as confirmed by the visiting and 15-minute trips coefficients.
- Perception of level of service measures altering mode share: when respondents report lower satisfaction of traffic levels, the relative utility of public transport drops significantly compared to the motorized baseline. This effect is the most prominent for commuting (β=-0.984), suggesting that mass transit is more penalized by congestion than private cars. For discretionary trips a perhaps counterintuitive result is obtained: in visiting (β=-0.366) and 15-minute (β=-0.348) trips a high traffic perception is associated with a reduction in utility for unmotorized modes compared to motorized ones. Interpretations on this

phenomenon could be oriented towards a reduced safety perception for pedestrians or various concerns for health.

- The controversial result regarding cleanliness: the "Clean_Opt" variable exhibits interesting outcomes for the commuting trip purpose. While low satisfaction with cleanliness translates into a penalization for public transport (β=-0.657), it conversely increases the likelihood of choosing active mobility (β=1.147). The first result is easily understandable because, when the streets are perceived as poorly maintained and unclean, the user typically prefers the private vehicle. Unlike public transport, which requires walking to transit stops and waiting in exposed public areas, the private car offers a protected environment that minimizes the user's direct physical exposure to street degradation. The second outcome is highly counterintuitive, but could hide behaviours that are not explainable with a simple MNL model. Rather than implying that users prefer dirty streets, the coefficient could serve as a proxy for other aspects not considered in the present model. This apparently anomalous result could justify using a more advanced model if the ICLV results prove to be satisfactory (4.2).

4.1.3 IIA Test Results

Following the theoretical considerations regarding the IIA (Independence of Irrelevant Alternatives) assumption (2.2.3), it is necessary to empirically validate this property through the Hausman-McFadden test. The base premise of the IIA assumption is that the relative probability of choosing between any two specific alternatives is completely independent of the presence or absence of other alternatives in the choice set. Consequently, if the IIA assumption holds, systematically removing a choice alternative from the estimation process should not significantly alter the parameter estimates (β) of the remaining alternatives.

To rigorously perform this test, the baseline Multinomial Logit model is compared to its restricted counterparts. Since the "Motorized" mode represents the reference in this analysis, the testing procedure involves estimating two distinct reduced models. The first is obtained by entirely removing the choice observations for the "Unmotorized" alternative and the second by removing the "Public Transport" one. Then, the Hausman-McFadden statistic compares the coefficients of these reduced databases with those of the full and unrestricted model. A statistically significant difference between the estimators would indicate a violation of the IIA property, suggesting a correlation among the unobserved utilities. The results of this comparison are presented in Table 19.

Table 19: Results of Hausman-McFadden Test (Source: own elaboration through SPSS)

Test Statistic	Removal of "Unmotorized"	Removal of "Public Transport"
H	-348.51	3.30
df	8	8
p-value	N.A.	0.91
Rejection of IIA	*No*	*No*

Given that the test statistics from the two restricted models fail to reject the null hypothesis, the IIA assumption is not violated. Consequently, the structural validity of the preliminary MNL model is confirmed.

4.2 Advanced Model Results: Integrated Choice and Latent Variable

4.2.1 Model Estimation and Goodness of Fit

Due to the complexity of the ICLV model, which requires the calculation of multiple integrals for the latent constructs, the four models were estimated using Maximum Simulated Likelihood (MSL) with 1000 Halton draws. This high number of iterations, obtained with the Monte-Carlo method, ensures the stability and precision of the estimators. Table 20 illustrates the goodness of fit of the models across the different trip purposes.

Table 20: Goodness of Fit Indicators for the ICLV Model (Source: own elaboration)

Metric	Commuting	Visiting	Shopping	15-min
Valid Observations (N)	700	883	995	1053
Number of Estimated Parameters (k)	32	32	32	32
Model Chi-Square (χ^2)	5203.59	6364.42	7139.21	7681.65
Null Log-Likelihood (LL(0))	-7774.94	-9852.44	-11172.38	-11875.35
Final Log-Likelihood (LL(β))	-5173.14	-6670.23	-7602.77	-8034.53
McFadden Pseudo R^2	0.335	0.323	0.320	0.323
McFadden Pseudo R^2 Adjusted	0.331	0.320	0.317	0.321

The overall performance of the estimated models demonstrates a strong goodness of fit. Across all four trip purposes, the final adjusted R^2 values are stable between 0.315 and 0.335. As established in discrete choice literature (McFadden, 1977), a Pseudo R^2 value between 0.2 and 0.4 represents a good behavioural fit, indicating that the models are successfully capturing a decent portion of the choice variance. Furthermore, despite the high number of estimated parameters ($k = 32$) required by the simultaneous estimation of the measurement and structural equations, the adjusted R^2 shows only a marginal penalization across the four trip purposes. Indeed, this indicator considers the number of variables evaluated in the model, reducing the overall fit if some parameters have a low Chi-Square. This minimal discrepancy between the two indicators confirms the accuracy of the specification, proving that the inclusion of both the latent and socio-demographic variables is highly optimized. Thus, it is possible to affirm that the model is free of overfitting issues.

When evaluating the ICLV performance against the base Multinomial Logit (MNL) model presented in section 4.1.2, a fundamental methodological distinction has to be made. A direct comparison using global fit indexes, such as the Final Log-Likelihood or the overall Chi-Square is structurally inappropriate. While the MNL maximizes exclusively the likelihood of the mode choice, the ICLV maximizes the likelihood of both the discrete choice and the multiple measurement equations derived from the Likert-scale indicators. Consequently, the high Log-Likelihood in absolute value observed in the ICLV framework does not indicate a degradation in model quality, but depends on the intrinsic nature and structure of the methodology.

The true empirical superiority of the ICLV approach over the base MNL is, instead, illustrated by the clear improvement in the explanatory power for non-mandatory trips. While the MNL model for commuting has a higher adjusted R^2 compared to the ICLV one, for shopping, visiting and 15-minute trips the value increases significantly, revealing the added value of latent constructs. Indeed, this result establishes that, for discretionary trips, the traditional socio-demographic variables are insufficient. The introduction of hidden psychological structures successfully captures an important layer of unexplained variance, bringing the goodness of fit of discretionary trips to the same level as mandatory commuting.

4.2.2 Measurement and Structural Model Results

Before the analysis of the mode choice model, it is necessary to evaluate the internal performance of the latent variables, considering the indicators' sign and significance and the regressors' explanatory power. All the results are presented in Table 21, where σ represents the standard deviation of the error terms, α is the intercept of the indicator equation and λ is the loading parameter linking a latent variable to the respective indicator.

Table 21: Estimation Results for the Latent Constructs: Structural and Measurement Equations across Trip Purposes (Source: own elaboration)

Parameter Name	Commuting	Shopping	Visiting	15min
LV_1 Regressors				
β_{Age,LV_1}	0.017***	0.020***	0.022***	0.020***
LV_1 Indicators				
σ_{living}	0.141	0.315***	0.347***	0.363***
α_{living}	-0.877***	-0.874***	-0.809***	-0.854***
λ_{living}	0.926***	0.863***	0.808***	0.851***
σ_{access}	0.916***	0.912***	0.933***	0.913***
α_{access}	-0.225***	-0.272***	-0.304***	-0.264***
λ_{access}	0.196***	0.255***	0.262***	0.265***
σ_{safety}	0.766***	0.790***	0.796***	0.796***
α_{safety}	-0.255***	-0.343***	-0.236***	-0.341***
λ_{safety}	0.403***	0.339***	0.271***	0.339***
LV_2 Indicators				
$\sigma_{traffic}$	0.846**	1.004***	0.747***	0.437***
$\alpha_{traffic}$	3.047***	3.121***	3.060***	3.129***
$\lambda_{traffic}$	0.853**	0.695***	0.952***	1.149***
$\sigma_{parking}$	1.009***	0.953***	1.009***	1.065***
$\alpha_{parking}$	2.422***	2.476***	2.495***	2.473***
$\lambda_{parking}$	0.472**	0.598***	0.454***	0.361***

*Note: ***: p<0.01, **: p<0.05, *: p<0.1*

Prior to the analysis of the LV's indicators, it is necessary to investigate the structural equations of the hidden constructs. While LV_2 does not have regressors due to the reasons explained in section 3.4.2.3, β_{Age,LV_1} exhibits a positive coefficient with a stable value across the four models. Ranging from 0.017 to 0.022, the demographic parameter influencing the "15-min City Attitude" construct indicates that older individuals tend to display a stronger alignment with the 15-minute city principles compared to younger age brackets. The ability to reach essential services, live in a rooted community and feel safe in the neighbourhood are factors that acquire importance as people age.

Regarding the measurement part, the results confirm the robustness of the hypothesised psychological constructs. For the first latent variable (LV_1), the model exhibits highly significant factor loadings for the indicators related to residential quality, accessibility and safety within a

15-minute radius. Specifically, the "λ_{living}" factor loading shows the strongest magnitude across all trip purposes, followed by "λ_{safety}" and "λ_{access}", which delineates a clear ranking across the indicators. Nevertheless, the strong statistical significance of the three indicators and their common positive sign demonstrates that the respondents' evaluations on the Likert scale are reliable manifestations of this underlying construct. The values deviations across the different trip purposes are not relevant, which signifies that the latent psychological attitude is independent from the trip context. It is another proof of the validity of the LV_1.

Similarly, the LV_2 capturing the "Perceived Easiness of Car Use" through traffic and parking sensitivity, is well defined by both its indicators. The satisfactions with traffic congestion and parking availability show positive and highly significant weights, confirming the coherence of the construct. On average, across the four trip purposes, the magnitude of "$\lambda_{traffic}$" is approximately twice as large as "$\lambda_{parking}$", suggesting that traffic congestion represents a much stronger psychological factor than parking availability.

4.2.3 Choice Model Results

The estimation of the discrete choice component within the ICLV framework confirms the expected behavioural dynamics regarding the traditional socio-economic attributes, which yields similar results as the MNL model. However, the Hybrid Model reveals interesting psychological hints through the inclusion of the latent constructs. Table 22 reports the parameter estimates grouped by their respective utility functions.

Table 22: Estimation Results of Choice Model (Source: own elaboration)

Variable Name	Description	Commuting	Shopping	Visiting	15min
Motorized	V_{Mot}				
$\beta_{Education\_Level\_Ordinal\_Opt,Mot}$	Education Level (Ref: No Tertiary Education)	0.787***	0.398**	0.504***	0.405**
$\beta_{MV\_Ordinal\_1,Mot}$	1 Motor Vehicle (Ref: No Motor Vehicle)	2.406***	2.266***	2.070***	1.483***
$\beta_{MV\_Ordinal\_2+,Mot}$	2 or more Motor Vehicles (Ref: No Motor Vehicle)	2.726***	2.521***	2.265***	1.679***
$\beta_{Driver\_License\_Binary,Mot}$	Driver License (Ref: No Driver License)	1.634***	0.946***	1.057***	0.334
Unmotorized	V_{unm}				
ASC_{Unm}	Intercept of V_{unm}	4.769***	5.633***	2.976***	3.627***
$\beta_{LV_1,Unm}$	LV_1: 15-min City Attitude	0.630***	0.360***	0.873***	0.533***
$\beta_{LV_2,Unm}$	LV_2: Perceived Easiness of Car Use	-0.418**	-0.386***	0.233**	0.193**

$\beta_{Age,Unm}$	Age	-0.160*	-0.148***	-0.072**	-0.038
$\beta_{Age^2,Unm}$	Age^2	0.00177*	0.00136***	0.00081**	0.00018
$\beta_{Dummy\_Povoa,Unm}$	Living in Povoa de Santa Iria (Ref: Living Outside)	-0.055	1.379***	1.653***	0.914***
Public Transport	V_{PT}				
ASC_{PT}	Intercept of V_{PT}	3.950**	1.677	3.420***	-0.185
$\beta_{Age,PT}$	Age	-0.156**	-0.101**	-0.188***	-0.096**
$\beta_{Age^2,PT}$	Age^2	0.00142	0.00094**	0.00166***	0.00088**
$\beta_{PT\_Card\_Binary,PT}$	Public Transport Card (Ref: No Transport Card)	3.862***	2.311***	2.715***	3.388***
$\beta_{Clean\_Opt,\,PT}$	Cleanliness of Streets (Ref: Low Satisfaction towards Cleanliness)	0.835**	0.106	0.862***	0.509*
$\beta_{Dummy\_Sobr,PT}$	Living in Sobralinho (Ref: Living Outside)	0.754	0.402	1.790***	-5.505***

*Note: ***: p<0.01, **: p<0.05, *: p<0.1*

4.2.3.1 Manifest Variables Analysis

The following comments can be made on the above reported result concerning manifest variables (sociodemographic and cleanliness) across the four considered models.

- **Education Level:** Concerning education level, distinguishing between tertiary education and secondary education or below, the results are expected and totally aligned with previous findings in the literature (see section 2.4.1). More specifically, $\beta_{Education\_Level\_Ordinal\_Opt,Mot}$ is positive and highly significant across the four models. This indicates that individuals with a higher educational background are more prone to use motorized modes, even for discretionary trips. This behaviour is explained by the fact that education often serves as a proxy for higher income and a greater value of time (Mincer, 1974), making private transport more accessible and attractive.

- **Motor Vehicles:** The reasoning behind the vehicle number variable is perhaps methodologically critical because, although the variable was perfectly scalar, it was decided to convert it to an ordinal one. Since performing this operation increases the computational weight of the model, it needs to be the result of a precise and rational modelling choice. Indeed, other than being rooted in transport modelling theory, the predilection for the ordinal variable was motivated by the observation of empirical results. Standard continuous specifications assume a linear relationship between utility and number of vehicles, implying that the gain obtained by acquiring the second vehicle is the same as getting the first. However, the marginal utility does not remain constant, as suggested by random utility theory (Ben-Akiva & Lerman, 1985), and is expected to decrease due to saturation effects. In fact, the first vehicle satisfies the primary mobility needs of the household, while the second one is utilized for secondary purposes or to provide flexibility, thus a discrete specification of the factor is justified (Fang, 2008).

The ordinal variable, divided into three classes (0=0 vehicles, 1=1 vehicle, 2=2 or more vehicles) exhibits a strong non-linearity. In the visiting model, the $\Delta\beta_{MV,Mot,Vis}$ between the first two categories (from zero to one vehicle) is 2.070, while between the second and the third (from one to two or more vehicles) is only 0.195. A similar pattern is observed also for the other trip purposes. Therefore, a linear specification would have forced the model to treat the second vehicle as equally valuable as the first, leading to an estimation bias.

This outcome could be explained by the concept of captivity. Indeed, the transition from zero to one vehicle represents a definite break in transport behaviour.

- Zero vehicles: individuals are often forced to use public transport or soft mobility, which make them "captive users".
- One vehicle or more: individuals start to have the choice on how to move, motorized modes become a viable alternative.

The option of creating a binary variable that only contemplates vehicle availability was critically examined and discarded due to the significance of owning a second vehicle and the dynamics of intra-household competition (Adjemian & Williams, 2009). In fact, if a family is formed by two adults but disposes of only one car, the two commuters cannot simultaneously attain their workplace. For this reason, vehicle availability is generally not an individual attribute, but a household resource, that serves all family members. Since work is the primary and the most prioritized activity, the marginal utility of the second car is expected to be higher in the commuting model than in the visiting one. The empirical results confirm this assumption, as the gap in question is 0.195 for the latter and 0.320 for the first. The second vehicle is, therefore, definitely more important for workers than for visitors.

- **Driver License:** Another fundamental observed attribute influencing the motorized utility is the possession of a driving license. For commuting trips ($\beta_{Driver\_License_{Binary},Mot,Com} = 1.634$) and standard discretionary activities, such as visiting ($\beta_{Driver\_Licens\ _{Binary},Mot,Vis} = 1.057$) and shopping ($\beta_{Driver\_Licen\ _{Binary},Mot,Shop} = 0.964$), this parameter is positive and highly significant ($p < 0.01$). Naturally, holding a license is a primary prerequisite that increases substantially the baseline utility of motorized mode and the resulting probability of choosing a private vehicle. However, an interesting behavioural insight emerges when analysing the 15-minute City scenario: in this specific context, the parameter drops in magnitude and completely loses its statistical significance ($p > 0.10$). This empirical outcome represents a clear validation of the 15-minute city paradigm. It suggests that, when amenities are accessible and activities are located nearby, the possession of a driving license ceases to be a determinant factor in mode choice. Indeed, since the structural advantage of licensed drivers is neutralized, active mobility becomes a competitive and natural choice regardless of the mobility assets owned by users (Scheiner, 2010).

- **Age:** Regarding the "Age" variable, the situation is more nuanced. It was chosen to adopt the quadratic approach described previously to prevent a potential erroneous linear relationship between endogenous and exogenous variables. The results, in fact, are different for each one of the modes, which reflects the evolving mobility needs across different life-stages:

  - Public Transport: For PT, the relationship with age generally exhibits a U-shaped curve, where both the linear and quadratic terms are significant (apart from $\beta_{Age^2,PT,Com}$ for commuting), with $\beta_{Age,PT} < 0$ and $\beta_{Age^2,PT} > 0$. This means that utility is the highest for young people and seniors reaching a low in middle adulthood. It is consistent with previous findings and also with common sense. In youth, car ownership and economic independence are relatively low and affect greatly the possibility of young people to use other transport means (captive users). The active adults are enjoying the peak of professional activity where the flexibility and speed of private car outweigh the importance of public transport lower cost, also for leisure and visiting purposes. For the elders, the recovery in PT use is explained by a reduced capability in driving and maybe in a lower time pressure due to retirement from professional activity. Furthermore, it is crucial to note a behavioural difference in the commuting model, where only the negative linear term is significant. This can be explained by the fact that commuting cease to be a determinant trip purpose after retirement age, neutralizing the increase of PT utility for mandatory trips.
  - Unmotorized: The model captures a similar U-shaped dynamic for commuting, shopping and visiting. The negative linear term coupled with the positive quadratic term shows a regression of unmotorized utility during middle age, with a resurgence for upper age brackets. Seniors tend to use active modes for leisure trips and to reach amenities and relatives in the neighbourhood. Conversely, in the

15-minute City scenario, neither age term is statistically significant ($p > 0.10$). This represents an interesting behaviour: a well-designed urban environment encourages active mobility unanimously, not considering the demographic differences between age groups.

- Motorized: Age parameters were omitted from the motorized utility to represent the baseline and for this reason both terms were set to 0. Because the utilities for active mobility and public transport significantly decrease as individuals age towards adulthood, the mathematical structure of the Logit model implies an automatic increase in the relative probability of choosing a private vehicle. Behaviourally, this illustrates that between the ages of 40 and 60, which corresponds to the career peak of a general worker, the private vehicle becomes the dominant alternative.

The previous explanations are supported graphically by Figure 9.

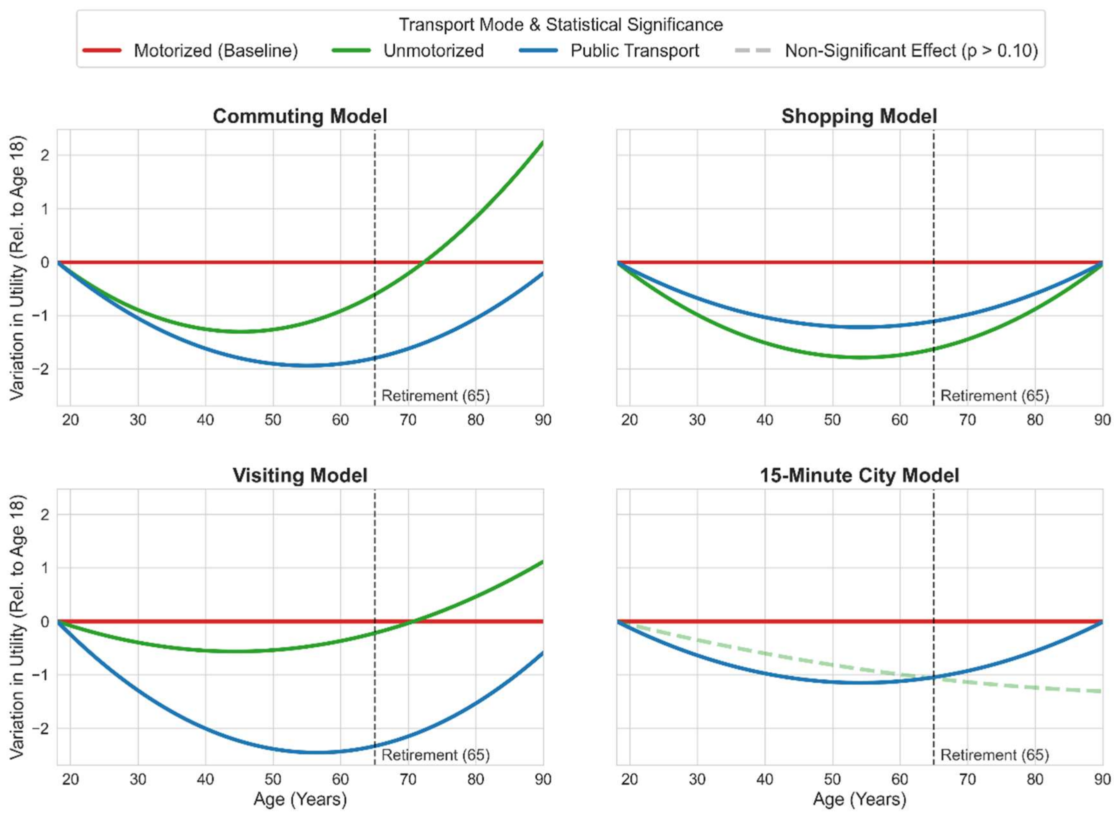

Figure 9: Marginal Utility Variation by Age across the Four Trip Purposes, Normalized at 18 Years Old (Source: own elaboration)

The graph illustrates the marginal utility variation in function of age. Values are normalized to highlights relative changes compared with the baseline set at 18 years old.

The U-shaped behaviour is particularly evident for commuting, shopping and visiting for both the alternative modes compared with the motorized baseline. The latter exhibits a growing trend compared to the other modes until approximately 50 years old depending on the purposes. At this point, it starts to lose attractivity, a trend that increases especially after retirement age. The dotted curve in the last model shows that age is not significant concerning unmotorized trips.

- **Spatial Bias:** To consider the unobserved spatial bias and the distinct structural layouts of the study area, specific location dummy variables were introduced for the residents of Póvoa de Santa Iria (applied to the unmotorized utility) and Sobralinho (applied to the public transport utility), leaving the motorized mode as the baseline.

  For Póvoa the results highlight a significant drive towards unmotorized mobility for the discretionary trip purposes. As mentioned in section 3.2.4, the location exhibits an urban layout that favours active mobility. However, the commuting coefficient is not significant ($p > 0.10$), indicating that mandatory work trips likely exceed walkable distance thresholds even in this favourable setting.

  Regarding Sobralinho, the coefficient shows a powerful positive propensity for PT in visiting trips ($\beta_{Dummy\_Sobr,PT,Vis} = 1.970, p < 0.01$), suggesting that residents rely heavily on transit connections for social trips. However, the parameter is statistically non-significant for commuting and shopping and exhibits a massive negative coefficient for 15-min trips ($\beta_{Dummy\_Sobr,PT,15} = -5.505, p < 0.01$). This anomaly likely captures the fact that the public transport lines are oriented towards suburban services, which are substantially useless for 15-min trips. Furthermore, a detailed analysis of the dataset reveals a case of perfect separation: zero observations were recorded for respondents residing in Sobralinho choosing public transport for the 15-minute trip scenario. This results in the least robust variable of the entire set, but it was retained to maintain consistency with the MNL model. It ensures a consistent comparative framework, allowing for an evaluation of the added value of latent variables without introducing bias in the model through the omission of certain parameters.

- **PT Card:** Concerning the impact of transport pass ownership across all models, where the coefficients range from 2.311 to 3.862, it is crucial to mention the issue of reverse causality. In this framework it can be formulated as follows:

  - Do I take the bus because I have a pass? ($Pass \rightarrow Choice$)
  - Do I buy the pass because I plan to take the bus? ($Choice \rightarrow Pass$)

It is possible to argue that there is an uncertain superposition between the two options presented above. This can be solved by considering a hierarchical decision-making process, where long-term and medium-term choices condition short-term mode choice decisions (Ben-Akiva & Lerman, 1985). Indeed, travel choices can happen in three distinct timeframes:

- Long-term: Residential or job location
- Medium-term: Mobility tools (buying a car, buying a PT pass, getting a license)
- Short-term: Daily mode choice (current model)

Since the model proposed focus on the lower tier of priorities, the causal link should be unidirectional, from the asset ownership to the modal choice. The process of acquiring a mobility tool is treated as a sunk cost, occurred in the past and not recoverable. This is especially true for discretionary trip purposes. In fact, people rarely buy a monthly pass specifically for recreation or to visit their friends occasionally. The decision is typically driven by mandatory trips and not leisure ones, making it effectively exogenous. Despite the validity of this hypothesis, concerning commuting, the risk of reverse causality is real. Indeed, comparing the coefficients between commuting and visiting models, there should be a noticeable increase towards the first one, which captures the endogeneity described above. $\Delta\beta_{PT\_Card\_Binary,PT} = \beta_{PT\_Card\_Binary,PT,Com} - \beta_{PT\_Card\_Binary,PT,Vis} = 1.147$. This implies that the structural strength of the pass is $e^{\Delta\beta_{PT\_Card\_Binary,PT}} = e^{1.147} = 3.149$ times larger in the commuting model than in the visiting one. It is a positive effect that could be explained also by the mandatory nature of the commuting trips. While visiting is a secondary activity, commuting is the action that justifies the expense of a transport pass, increasing the utility weight. Furthermore, the act of buying a PT card contributes to break down the marginal cost of travelling, incentivizing the user to utilize it as much as possible.

- **Cleanliness:** The results indicate that cleanliness is a significant factor for public transport utility, but not for unmotorized modes. This perhaps counterintuitive result finds support in transport psychology literature. Indeed, for PT, cleanliness acts as an important component for service quality and personal security (Eboli & Mazzulla, 2007), particularly during waiting times, where users are static and vulnerable to environmental degradation. The coefficient analysis clearly confirms this dynamic: the impact of cleanliness on transit utility is positive and highly significant for both visiting ($\beta_{Clean\_Opt,\,PT,Vis} = 0.862, p < 0.01$) and commuting ($\beta_{Clean\_Opt,\,PT,Com} = 0.835, p < 0.05$). On the other hand, regarding active mobility, in the “Hierarchy of Walking Needs” (Alfonzo, 2005), the author suggests that "pleasurability" (which includes cleanliness) is a secondary need that becomes relevant only if other basic requirements like feasibility, accessibility and traffic safety are met. In a 15-minute city context, the decision to walk is even more linked to these aspects (which are incorporated into LV_1), making factors like cleanliness secondary and statistically non-significant in this sample.

4.2.3.2 Latent Variables Analysis

- **LV_1: "Attitude towards 15-min City":** This latent construct represents the core behavioural concept introduced in the ICLV model. Its statistical significance and mathematical coherence could alone justify the use of a hybrid specification instead of a simple MNL one. The parameter capturing the impact of LV_1 on the unmotorized utility ($\beta_{LV_1,Unm}$) exerts a strongly positive and highly significant effect across all four evaluated trip purposes. Although a positive impact is observed for commuting and shopping, its influence peaks in the visiting trip purpose ($\beta_{LV_1,Unm,Vis} = 0.872, p < 0.01$) and is remarkable also in the 15-min model ($\beta_{LV_1,Unm,15} = 0.533, p < 0.01$).

  These results confirm the previous hypothesis concerning the psychological evaluation of the urban space (2.1.2). The indicators composing LV_1 highlight some of the fundamental requirements that must be guaranteed for an individual to choose active mobility. Liveability, access to amenities and safety perception are the core indicators pushing citizens to adopt greener modes, unlike other often mentioned aspects that do not appear to be significant and will be discussed later (5.3). Without the current framework, the preferences that lead users towards active mobility would not have been directly measured, being absorbed by the error term. This limitation generates unexplained variance and possible biases in simpler models, as seen in the MNL, which was yielding some inaccurate and apparently inexplicable results (4.1.2.2).

- **LV_2: "Perceived Easiness of Car Use":** The second latent variable reveals an interesting behavioural dynamic through a clear sign reversal in its parameters, depending on the trip purpose. For mandatory trips and maintenance activities, the impact of this attitude on the utility of unmotorized modes is logically negative and significant: $\beta_{LV_2,Unm,Com} = -0.418, p < 0.05$ for commuting and $\beta_{LV_2,Unm,Shop} = -0.386, p < 0.05$ for shopping. In these contexts, a user who places a high value on the comfort of a private vehicle noticeably penalizes active mobility. The motorized alternative is preferred for reasons of pure convenience, such as the necessity of carrying groceries or heavy loads for shopping or the desire to minimize exposure to congestion on long commuting trips.

However, the results manifest a fundamental behavioural shift for discretionary and proximity-based trips: in visiting ($\beta_{LV_2,Unm,Vis} = 0.233, p < 0.05$) and 15-minute city ($\beta_{LV_2,Unm,15} = 0.192, p < 0.05$) trips, the coefficients are surprisingly positive, albeit being modest in magnitude. Since a high value of LV_2 corresponds to a strong satisfaction towards traffic levels and parking availability, this outcome reveals a peculiar dynamic. Indeed, a neighbourhood that has a lower level of vehicular flow is essentially a quieter and safer environment. For short-distance trips and local visits, this characteristic helps to mitigate some of the limitations to walking, such as noise, pollution and pedestrian hazard. Consequently, an urban space that is free of traffic not only satisfies drivers, but simultaneously provides a safe walking and cycling environment and encourages residents to choose active modes to reach nearby destinations.

5. Conclusions

5.1 General Conclusions

This thesis explored the 15-minute city concept by integrating transport mode choice and trip purpose influence, offering a valuable behavioural perspective on the subject. The combination of traditional discrete choice modelling and Integrated Choice and Latent Variable (ICLV) approach aims to justify the better explanatory power of the second, through the hidden behavioural dimensions influencing travel decisions.

The results obtained from the empirical analysis highlight a subtle distinction between commuting and non-mandatory trips in terms of behavioural patterns. Indeed, commuting trips appear to be primarily driven by structural constraints, such as car, driver license or transport card ownership. These elements significantly reduce the flexibility of individuals, leading to more stable and habitual modal choices. Furthermore, the coefficients of the previously mentioned mobility assets are frequently the largest across the four models, demonstrating that their effect is preponderant.

On the contrary, non-mandatory trips, like visiting and the 15-minute spatial filter, exhibit a higher degree of flexibility. In this case, the decision-making process is not as much determined by objective transport attributes, but is also influenced by subjective and attitudinal factors. In particular, the inclusion of latent variables in the modelling framework allows to capture the role of perceptions and preferences. Indeed, the results indicate that neighbourhood awareness and car comfort have a statistically significant effect on mode choice, suggesting that travel behaviour cannot be totally explained by observable variables alone.

Overall, the outcomes of the various models support the idea that policies promoting sustainable mobility and 15-minute city urban planning should not rely exclusively on infrastructural interventions. Neighbourhood perception and access to amenities must, therefore, be considered in the design of transport and urban policies. Moreover, these results highlight the extreme importance of land use in shaping the cities of the future.

5.2 Added Value of the Hybrid Choice Model

The comparison between the Multinomial Logit (MNL) model and the Integrated Choice and Latent Variable (ICLV) model provides useful insights concerning the role of psychological factors in the decision-making process. While the MNL model represents a solid baseline specification based on observable variables, it does not allow to explicitly capture the underlying behavioural mechanisms. On the other hand, the ICLV framework makes it possible to include latent constructs, thus enabling a more detailed representation of individual preferences.

In particular, the ICLV model allows to decompose the effects of socio-demographic variables into direct and indirect components. Indeed, some observable characteristics are influencing modal choice not only directly, but also indirectly through their impact on attitudes and perceptions. This aspect is particularly relevant in the context of this study, where a regressor such as "Age" appears in the equation of both mode choice and LV_1. In the study context, where subjective factors are expected to be relevant, this is a crucial enhancement in the model prediction power. The significance of the two LVs confirms the importance of integrating psychological dimensions in transport modelling.

5.3 Policy Implications

The results are suggesting that transport policies should consider the heterogeneity of travel behaviour across trip purposes, since adopting a uniform strategy could be less effective in promoting sustainable mobility. Both the observable independent variables and the two latent constructs prove to have different magnitudes depending on the reason that drives people to move. Consequently, strategies designed to discourage car usage should not be applied equally to long-distance commuting and local discretionary trips.

Regarding commuting, policy interventions should focus on improving the attractiveness of public transport systems. Indeed, the high impact of transport card possession on PT utility confirms the importance of pass costs and incentives. Nevertheless, recent reductions in seasonal ticket prices suggest that further increases in public transport demand are more likely to be achieved through improvements in service quality and supply. Furthermore, the significance of indicators such as the perceived cleanliness of the streets highlights that financial incentives alone are insufficient. To effectively compete with the comfort of private cars, the passenger has to experience a pleasant journey fostered by the high-quality service of the public transit.

Conversely, non-mandatory trips offer a greater potential for policies targeting the change of attitudes captured by the analysed latent constructs. In particular, the latent variable related to the attitude towards the 15-minute city (LV_1), which is influenced by perceived safety, accessibility to amenities and overall liveability of the neighbourhood, shows a significant impact on the adoption of active modes. This suggests that improving the quality of the local urban environment can directly contribute to a modal shift for short-distance trips. Therefore, policies focussing on the enhancement of pedestrian safety, availability of local amenities and overall attractiveness of neighbourhoods can play a key role in encouraging walking and cycling.

Furthermore, the direct comparison between the two latent variables (LV_1 and LV_3) concerning the urban environment produces interesting insights, recalling that the latter was not found significant. According to the coefficients, neighbourhood quality and perceived safety have a stronger impact on modal choice compared to elements related to urban decorum and pedestrian infrastructure. Aesthetic factors and sidewalk facilities appear to have a limited role in convincing people to adopt active mobility.

At the same time, the latent variable representing the "Perceived Easiness of Car Use" (LV_2), which depends on perception of traffic conditions and parking availability, highlights the importance of perceived convenience in the use of the car. The results suggest that when the use of private transport is seen as easy and comfortable, individuals are more likely to rely on it, but only for maintenance trips (commuting and shopping). Consequently, policies aimed at reducing parking spots and limit attractiveness of private vehicles could be ineffective, not necessarily increasing utility of active modes for discretionary trips.

5.4 Limitations of the Present Case Study

Specific limitations due to various constraints in the dataset or sample methods should be acknowledged:

- The dataset does not include information on travel time and cost, which are traditionally among the most relevant independent variables in transport mode choice modelling. Their absence may limit the explanatory power of the estimated models and partially affect the interpretation of some behavioural effects.
- The location variable refers to the interview position rather than the actual residential location of respondents. Therefore, the spatial interpretation of the results should be treated with caution, since the collected information might not completely reflect the environmental conditions experienced by individuals in their daily lives.
- The sampling process was not fully randomized. Although the survey adopted a quota-based approach aimed at approximating the socio-demographic structure of the population (age and gender), the sample remains non-probabilistic. Therefore, the findings should not be interpreted as statistically representative of the local population, but as a reasonable approximation.

- The study focuses on the Lisbon metropolitan area and, more specifically, on a suburban context characterised by its own spatial and mobility dynamics. For this reason, the transferability of the results to different urban environments could be limited.

5.5 Further Research

Additional research could help to address the issues mentioned in the previous section. Specifically, the main improvement would be achieved through the inclusion of additional level of service variables, such as cost and time of the different modal alternatives. More complete data about service reliability and income could also be valuable. In summary, a sample more oriented towards mode choice workflow might produce a better overall model estimation.

Apart from the methodological aspects, further research could investigate more deeply the interaction between behavioural attitudes and urban mobility policies. The findings of this study suggest that improving urban decorum and pedestrian paths might not be sufficient to substantially modify travel behaviour, especially for more constrained trip purposes such as commuting. Subsequent analyses could, therefore, investigate if a focused infrastructural intervention is actually shifting mode choice towards sustainable mobility. A "before/after" comparison could yield valuable indirect insights about the role of attitudes towards 15-minute city contexts.

Lastly, particular attention could also be addressed to the effects of emerging changes in commuting habits, including hybrid working patterns such as telework. These transformations could progressively alter the relationship between trip purpose, urban proximity and transport mode choice, potentially reshaping the effectiveness of the 15-minute city model in suburban environments.

References

Adamidis et al. (2025). Mode choices under car-reducing scenarios: Measurable factors and latent attitudes. *Travel Behaviour and Society*. doi:https://doi.org/10.1016/j.tbs.2025.101001

Adjemian, M., & Williams, J. (2009). Using census aggregates to proxy for household characteristics: an application to vehicle ownership. *Transportation*. doi:https://doi.org/10.1007/s11116-009-9191-2

Alfonzo, M. (2005). To Walk or Not to Walk? The Hierarchy of Walking Needs. *Environment and Behavior*. doi:https://doi.org/10.1177/0013916504274016

Arias-Molinares et al. (2025). Exploring the 15-Minute City concept for the urban outskirts: a systematic literature review. *European Transport Research Review*. doi:https://doi.org/10.1186/s12544-025-00743-8

Ben-Akiva, M., & Lerman, S. (1985). *Discrete choice analysis: Theory and application to travel demand.* MIT Press. doi:ISBN: 978-0-262-02217-0

Bhagat-Conway et al. (2022). Subjective variables in travel behavior models: a critical review and Standardized Transport Attitude Measurement Protocol (STAMP). *Transportation*. doi:https://doi.org/10.1007/s11116-022-10323-7

Bhat, C. R. (2001). Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research Part B: Methodological*. doi:https://doi.org/10.1016/S0191-2615(00)00014-X

Bierlaire, M. (2003). BIOGEME: A free package for the estimation of discrete choice models. *Proceedings of the 3rd Swiss transportation research conference*. doi:EPFL: 4d8f74a8-0b6e-4d15-b878-8e4377f3d24c

Bouscasse, H. (2018). Integrated choice and latent variable models: A literature review on mode choice. doi:https://hal.science/hal-01795630v1

Carvalho et al. (2025). Assessing the readiness for 15-minute cities: a literature review on performance metrics and implementation challenges worldwide. *Transport Reviews*. doi:https://doi.org/10.1080/01441647.2025.2513530

Chen, J., & Li, S. (2017). Mode Choice Model for Public Transport with Categorized Latent Variables. *Mathematical Problems in Engineering*. doi:https://doi.org/10.1155/2017/7861945

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences.* Lawrence Erlbaum Associates. doi:https://doi.org/10.4324/9780203771587

Colaço, R., & de Abreu e Silva, J. (2025). Does the 15-minute City Promote Sustainable Travel? *Networks and Spatial Economics*. doi:https://doi.org/10.1007/s11067-025-09670-6

Crane, R. (2007). Is There a Quiet Revolution in Women's Travel? Revisiting the Gender Gap in Commuting. *Journal of the American Planning Association*. doi:https://doi.org/10.1080/01944360708977979

Dargay, J. M. (2001). The effect of income on car ownership: evidence of asymmetry. *Transportation Research Part A: Policy and Practice*. doi:https://doi.org/10.1016/S0965-8564(00)00018-5

Dong et al. (2025). Geographic Scale Matters in Analyzing the Effects of the Built Environment on Choice of Travel Modes: A Case Study of Grocery Shopping Trips in Salt Lake County, USA. *Urban Science*. doi:https://doi.org/10.3390/urbansci9080307

Eboli, L., & Mazzulla, G. (2007). Service Quality Attributes Affecting Customer Satisfaction for Bus Transit. *Journal of Public Transportation*. doi:http://doi.org/10.5038/2375-0901.10.3.2

Elldér, E. (2024). The 15-minute city dilemma? Balancing local accessibility and gentrification in Gothenburg, Sweden. *Transportation Research Part D: Transport and Environment*. doi:https://doi.org/10.1016/j.trd.2024.104360

Ewing, R., & Cervero, R. (2010). Travel and the Built Environment: A Meta-Analysis. *Journal of the American Planning Association* . doi:https://doi.org/10.1080/01944361003766766

Fale et al. (2025). Systematic Review of Transportation Choice Modeling. *Applied Sciences*. doi:https://doi.org/10.3390/app15179235

Fang et al. (2014). Understanding Attitudes towards Proenvironmental Travel: An Empirical Study from Tangshan City in China. *Computational Intelligence and Neuroscience*. doi:https://doi.org/10.1155/2014/963683

Fang, H. A. (2008). A discrete–continuous model of households' vehicle choice and usage, with an application to the effects of residential density. *Transportation Research Part B: Methodological*. doi:https://doi.org/10.1016/j.trb.2008.01.004

Fernández Núñez et al. (2024). Car-use reduction in 15-Minute Cities. A matter of modal shift or shorter travel distances? *Journal of Urban Mobility*. doi:https://doi.org/10.1016/j.urbmob.2024.100093

Fernández-Heredia et al. (2014). Understanding cyclists' perceptions, keys for a successful bicycle promotion. *Transportation Research Part A*. doi:https://doi.org/10.1016/j.tra.2014.02.013

Frank et al. (2000). Linking land use with household vehicle emissions. *Transportation Research Part D: Transport and Environment*. doi:https://doi.org/10.1016/S1361-9209(99)00032-2

Galobardes et al. (2006). Indicators of socioeconomic position. *Journal of Epidemiology & Community Health*. doi:https://doi.org/10.1136/jech.2004.023531

George et al. (2025). Spatial advantages of highly educated individuals in Germany: Is sustainable mobility an expression of privilege? *Cities*. doi:https://doi.org/10.1016/j.cities.2024.105507

Gim, T. H. (2017). Full Random Coefficients Multilevel Modeling of the Relationship between Land Use and Trip Time on Weekdays and Weekends. *Sustainability*. doi:https://doi.org/10.3390/su9101824

Gould, K., & Lewis, T. (2016). *Green Gentrification.* Yale University Press. doi:ISBN: 9781138309135

Hanson, S., & Pratt, G. (1995). *Gender, Work, and Space.* Routledge. doi:ISBN: 9780415099417

Hausman, J., & McFadden, D. (1984). Specification tests for the multinomial logit model. *Econometrica*. doi:https://doi.org/10.2307/1910997

Jameel, A. K., & Abdulhussein, A. A. (2024). Incorporating Latent Variables into Active Transportation Mode Choice: A review. *Al-Rafidain Journal of Engineering Sciences*. doi:https://doi.org/10.61268/vhyhs009

Jeong et al. (2022). Travel mode choice as a representation of travel utility: A multilevel approach reflecting the hierarchical structure of trip, individual, and neighborhood characteristics. *Papers in Regional Science*. doi:https://doi.org/10.1111/pirs.12665

Johannson et al. (2006). The effects of attitudes and personality traits on mode choice. *Transportation Research Part A: Policy and Practice*. doi:https://doi.org/10.1016/j.tra.2005.09.001

Kaiser, H. F., & Rice, J. (1974). Little Jiffy, Mark Iv. *Education and Psychological Measurement*. doi:https://doi.org/10.1177/001316447403400115

Kalantari et al. (2025). Machine Learning in Mode Choice Prediction as Part of MPOs' Regional Travel Demand Models: Is It Time for Change? *Sustainability*. doi:https://doi.org/10.3390/su17083580

Kamargianni, M., & Polydoropoulou, A. (2013). Hybrid Choice Model to Investigate Effects of Teenagers' Attitudes toward Walking and Cycling on Mode Choice Behavior. *Transportation Research Record*. doi:https://doi.org/10.3141/2382-17

Khavarian-Garmsir et al. (2023). From Garden City to 15-Minute City: A Historical Perspective and Critical Assessment. *Land*. doi:https://doi.org/10.3390/land12020512

Kim et al. (2020). An Empirical Analysis for Mode Choice in a Short-Distance Trip with Personal Rapid Transit. *Journal of Advanced Transportation*. doi:https://doi.org/10.1155/2020/7436710Digital Object Identifier (DOI)

Kim, J., & Li, M. (2023). Campus commute mode choice in a college town: An application of the integrated choice and latent variable (ICLV) model. *Travel Behaviour and Society*. doi:https://doi.org/10.1016/j.tbs.2022.10.002

Loa, P., & Habib, K. N. (2023). Identifying the Determinants of Anticipated Post-Pandemic Mode Choices in the Greater Toronto Area: A Stated Preference Study. *Transportation Research Record: Journal of the Transportation Research Board*. doi:https://doi.org/10.1177/03611981221145133

Loukaitou-Sideris, A. (2006). Is it Safe to Walk? Neighborhood Safety and Security Considerations and Their Effects on Walking. *Journal of Planning Literature*. doi:https://doi.org/10.1177/0885412205282770

Loukaitou-Sideris, A. (2014). Fear and safety in transit environments from the women's perspective. *Security journal*. doi:https://doi.org/10.1057/sj.2014.9

Lu, H., & Gan, H. (2024). Unraveling the Influence of Perceived Built Environment on Commute Mode Choice Based on Hybrid Choice Model. *Applied Sciences*. doi:https://doi.org/10.3390/app14177921

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis.* Wiley. doi:ISBN: 978-0-486-44136-8

Mariel et al. (2024). An assessment of the current use of hybrid choice models in environmental economics, and considerations for future applications. *Journal of Choice Modelling*. doi:https://doi.org/10.1016/j.jocm.2024.100520

Martínez, L. M., & Viegas, J. M. (2013). A new approach to modelling distance-decay functions for accessibility assessment in transport studies. *Journal of Transport Geography*. doi:https://doi.org/10.1016/j.jtrangeo.2012.08.018

McFadden, D. (1977). Quantitative Methods for Analyzing Travel Behaviour of Individuals: Some Recent Developments. *Cowles Foundation Discussion Paper No. 474*. doi:https://doi.org/10.4324/9781003156055-18

Mincer, J. (1974). *Schooling, experience, and earnings.* Columbia University Press. doi:https://doi.org/10.1086/260336

Moreno et al. (2021). Introducing the "15-Minute City": Sustainability, Resilience and Place Identity in Future Post-Pandemic Cities. *Smart Cities*. doi:https://doi.org/10.3390/smartcities4010006

Moreno, C. (2016). La ville du quart d'heure: pour un nouveau chrono-urbanisme ! *La Tribune*. doi:Website: https://www.latribune.fr/regions/smart-cities/la-tribune-de-carlos-moreno/la-ville-du-quart-d-heure-pour-un-nouveau-chrono-urbanisme-604358.html

Mosquera et al. (2023). 15-min Cities: the Potential of a Medium-Sized Polycentric Latin American City. *Journal of Urban Health*. doi:https://doi.org/10.1007/s11524-023-00749-4

Murooka et al. (2025). Differences in choice of transportation mode by residents of 15-minute cities according to destination amenities: Break away from car dominant city. *Transportation Research Part A: Policy and Practice*. doi:https://doi.org/10.1016/j.tra.2025.104452

Negm et al. (2023). Exploring the X-Minute City by Travel Purpose in Montréal, Canada. *Transport Findings*. doi:https://doi.org/10.32866/001c.77506

Ortúzar, J. d., & Willemsen, L. G. (2011). *Modelling transport.* John Wiley & Sons. doi:https://doi.org/10.1002/9781119993308

Ranjan, R., & Sinha, S. (2025). A systematic review of mode choice behavior in urban transportation with emphasis on individual preferences and influencing factor. *Discover Cities*. doi:https://doi.org/10.1007/s44327-025-00138-3

Rhemtulla et al. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*. doi:https://doi.org/10.1037/a0029315

Sándor, Z., & Train, K. (2004). Quasi-random simulation of discrete choice models. *Transportation Research Part B: Methodological*. doi:https://doi.org/10.1016/S0191-2615(03)00014-6

Scheepers et al. (2014). How Combined Trip Purposes Are Associated with Transport Choice for Short Distance Trips. Results from a Cross-Sectional Study in the Netherlands. *PLoS One*. doi:https://doi.org/10.1371/journal.pone.0114797

Scheiner, J. (2010). Interrelations between travel mode choice and trip distance: trends in Germany 1976–2002. *Journal of Transport Geography*. doi:https://doi.org/10.1016/j.jtrangeo.2009.01.001

Shrestha, N. (2021). Factor Analysis as a Tool for Survey Analysis. *American Journal of Applied Mathematics and Statistics*. doi:https//doi.org/10.12691/ajams-9-1-2

Taber, K. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education*. doi:https://doi.org/10.1007/s11165-016-9602-2

Toorzani, A., & Rassafi, A. (2022). Pro-environmental attitude and adherence to a travel mode in an integrated choice and latent variable (ICLV) model: results from a revealed preference survey. *International Journal of Civil Engineering*. doi:https://doi.org/10.1007/s40999-022-00757-6

Train, K. (2009). *Discrete Choice Methods with Simulation.* Cambridge University Press. doi:https://doi.org/10.1017/CBO9780511753930

Tyrinopoulos, Y., & Antoniou, C. (2013). Factors affecting modal choice in urban mobility. *European Transport Research Review*. doi:https://doi.org/10.1007/s12544-012-0088-3

Upahita et al. (2022). The Influence of Trip Purpose on the Mode Choice Between High-Speed. *RSF Conference Series: Engineering and Technology*. doi:https://doi.org/10.31098/cset.v2i2.576

Verhoeven et al. (2016). Psychosocial and Environmental Correlates of Walking, Cycling, Public Transport and Passive Transport to Various Destinations in Flemish Older Adolescents. *PLoS ONE*. doi:https://doi.org/10.1371/journal.pone.0147128

Vij et al. (2013). Incorporating the influence of latent modal preferences on travel mode choice behavior. *Transportation Research Part A: Policy and Practice*. doi:https://doi.org/10.1016/j.tra.2013.07.008

Vij, A., & Walker, J. (2016). How, when and why integrated choice and latent variable models are latently useful. *Transportation Research Part B: Methodological*. doi:https://doi.org/10.1016/j.trb.2016.04.021

A1: Database Codebook

Variable name	General description	From Survey/Derived	Survey code	Survey description/Derivation method
Socio-Demographic and Residential Attributes				
Age	Age	Survey	AGE	What is your age?
Age_Ordinal	Ordinal value for "Age" variable	Derived	N.A.	0=Young<30, 1=Active<65, 2=Elders≥65
Age_Ordinal_Opt	Recoding of the previous variable into fewer categories	Derived	N.A.	1 if Elders (>65), 0 otherwise
Age_sq	Age^2	Derived	N.A.	Age^2
Dummy_Povoa	Binary value for Povoa de Santa Iria location	Derived	N.A.	1 if interview was done in Povoa de Santa Iria, 0 otherwise
Dummy_Sobr	Binary value for Sobralinho location	Derived	N.A.	1 if interview was done in Sobralinho, 0 otherwise
Education_Level	Education condition	Survey	23_EDUCATION	What is the highest level of education you have completed?
Education_Level_Ordinal	Ordinal education level	Derived	N.A.	Dividing Education_Level in 7 categories, from 0 to 6
Education_Level_Ordinal_Opt	Recoding of the previous variable into fewer categories	Derived	N.A.	Dividing Education_Level in 2 categories (Non-tertiary/Tertiary)
Gender	Gender	Survey	GENDER	What is your gender?
Gender_Dummy	Binary value for "Gender" variable	Derived	N.A.	0 if Male, 1 otherwise
Health	Health condition	Survey	19_HEALTH	How would you rate your health?
Health_Ordinal	Ordinal health condition	Derived	N.A.	Dividing Health in 5 categories, from 0 to 4
HH_Adults	Number of adults in household	Survey	16_YEARS_18_64	Not counting yourself, how many people make up your household at each of these ages? Adults (18 to 64 years)
HH_Babies	Number of babies in household	Survey	16_YEARS_0_9	Not counting yourself, how many people make up your household at each of these ages? Children (0 to 9 years)
HH_Elders	Number of elders in household	Survey	16_YEARS_65	Not counting yourself, how many people make up your household at each of these ages? Elderly (65 or more years)
HH_Teenagers	Number of teens in household	Survey	16_YEARS_10_17	Not counting yourself, how many people make up your household at each of these ages? Teenagers (10 to 17 years)
HH_Total	Number of people in household	Derived	N.A.	HH_Babies+HH_Teenagers+HH_Adults+HH_Elders+1
ID	ID	Survey	_id	_id
Income_Missing	Missing income value	Derived	N.A.	If Income_Range = N.A.
Income_Range	Income range	Survey	27_INCOME	What is your household's monthly gross income (before taxes)?

Income_Range_Avg	Average value of the income interval	Derived	N.A.	Making Income_Range scalar
Income_Range_Avg_k	Average value of the income interval in thousands	Derived	N.A.	Income_Range_Avg/1000
Income_Range_Ordinal	Ordinal value of income	Derived	N.A.	Dividing Income_Range in 3 categories of income (Low<1000<Medium<1750<High)
Internet_Use	Internet use	Survey	26_INTERNET	Do you have internet access via smartphone or computer at home?
Internet_Use_Binary	Binary coding (Yes/No) of the previous variable	Derived	N.A.	1 if Yes, 0 otherwise
Local	Municipality in which the interview was done	Survey	Local	Survey location
Local_Opt	Correcting "Local" variable	Derived	N.A.	Uniformed "2" and "3" in one category, same location
Residence_Type	Residence type	Survey	14_TYPERESIDENCE	What type of residence do you live in?
Residence_Type_Binary	Binary coding (Apartment/house) of the previous variable	Derived	N.A.	0 if Apartment, 1 otherwise
Residence_Years	Years of permanence in the same house	Survey	13_YEARSRESIDENCE	How many years have you lived in this residence?
Work	Work condition	Survey	20_EMPLOY	What is your current employment status?
Mobility Assets and Travel Behaviour				
Avg_Trip_Time	Average trip time range for commuting	Survey	22_MEANTIME	On average, how long does your commute to work and/or school take?
Avg_Trip_Time_Ordinal	Ordinal average trip time for commuting	Derived	N.A.	Dividing Avg_Trip_Time in 6 categories, from 0 to 5
Avg_Trip_Time_Ordinal_Opt	Recoding of the previous variable into fewer categories	Derived	N.A.	Dividing Avg_Trip_Time in 3 categories (Low<15min<Medium<30min<High)
Avg_Trip_Time_Scale	Average value of trip time interval	Derived	N.A.	Making Avg_Trip_Time scalar
Bikes	Number of Bikes	Survey	17_BIKE	How many vehicles does your household own? Bikes
Bus_Access	Satisfaction with bus ease of walking access	Survey	2_BUS_ACESS	Considering a scale of 1 to 5 where 1 means you are very dissatisfied and 5 – very satisfied, how satisfied are you regarding the ease of bus access
Bus_Inclination	Average between bus quantity and access	Derived	N.A.	(Bus_Quantity+Bus_Access)/2
Bus_Quantity	Satisfaction with bus quantity	Survey	2_BUS_QUANT	Considering a scale of 1 to 5 where 1 means you are very dissatisfied and 5 – very satisfied, how satisfied are you regarding the quantity of bus
Bus_Quantity_Opt1	Recoding "Bus_Quantity" variable into fewer categories	Derived	N.A.	0 if <3, 1 if 3, 2 if >3
Bus_Quantity_Opt2	Recoding "Bus_Quantity" variable into fewer categories	Derived	N.A.	0 if <4, 1 otherwise
Bus_Use	Frequency of bus use	Survey	9_BUS	How often do you visit/use the following amenities: Public transport stations and stops
Bus_Use_Ordinal	Ordinal frequency of bus use	Derived	N.A.	Transfomed the above frequency in an ordinal scale from 0 to 5
Driver_License	Driver License	Survey	24_DRIVER	Do you have a driver's license?
Driver_License_Binary	Binary coding (Yes/No) of the previous variable	Derived	N.A.	1 if Yes, 0 otherwise

Motor_Vehicles	Number of MV	Derived	N.A.	Cars+Motorcycles (17_AUTO+17_MOTO)
MV_Ordinal	Ordinal value for "Motor_Vehicles" variable	Derived	N.A.	0=0 veh, 1=1 veh, 2=2+ veh
MV_Ordinal_Opt	Recoding "MV_Ordinal" variable into fewer categories	Derived	N.A.	1 if have vehicles, 0 otherwise
PT_Card	PT Card	Survey	25_TRANSPORTCARD	Do you have a public transport pass (Navegante)?
PT_Card_Binary	Binary coding (Yes/No) of the previous variable	Derived	N.A.	1 if Yes, 0 otherwise
Mode Choice				
15min_Missing	Missing mode for 15-min trips	Derived	N.A.	0 if missing value, 1 otherwise
Commuting_Missing	Missing mode for commuting trip purpose	Derived	N.A.	0 if missing value, 1 otherwise
Leisure_Missing	Missing mode for leisure trip purpose	Derived	N.A.	0 if missing value, 1 otherwise
Mode_15mincity	Used mode for trips in 15-min city	Survey	11_MODAL	How do you usually travel in the area where you reside (within a radius of 800 meters, or approximately 15 minutes on foot around your residence)?
Mode_15mincity_Opt	Generalization of used mode for trips in 15-min city	Derived	N.A.	Regrouping Mode_15mincity in three macro-categories: Motorized, Unmotorized, PT
Mode_commuting	Used mode for commuting trip purpose	Survey	21_TRANSPORT	Last week, what was the mode of transport you used most to go to work and/or school?
Mode_commuting_Opt	Generalization of used mode for commuting trip purpose	Derived	N.A.	Regrouping Mode_commuting in three macro-categories: Motorized, Unmotorized, PT
Mode_leisure	Used mode for leisure trip purpose	Survey	10_LEISURE	Last week, what was the mode of transport you used most for the following trip purpose? For leisure (playing sports, having a coffee, walking in the park, going to the cinema)
Mode_leisure_Opt	Generalization of used mode for leisure trip purpose	Derived	N.A.	Regrouping Mode_leisure in three macro-categories: Motorized, Unmotorized, PT
Mode_shopping	Used mode for shopping trip purpose	Survey	10_SHOP	Last week, what was the mode of transport you used most for the following trip purpose? Shopping – food products, grocery, supermarket
Mode_shopping_Opt	Generalization of used mode for shopping trip purpose	Derived	N.A.	Regrouping Mode_shopping in three macro-categories: Motorized, Unmotorized, PT
Mode_visiting	Used mode for visiting trip purpose	Survey	10_VISITING	Last week, what was the mode of transport you used most for the following trip purpose? Visiting or being with friends
Mode_visiting_Opt	Generalization of used mode for visiting trip purpose	Derived	N.A.	Regrouping Mode_visiting in three macro-categories: Motorized, Unmotorized, PT
Multimodal_Binary	The user is not using the same mode for all trip purposes	Derived	N.A.	1 if the user is using different modes across trip purposes, 0 otherwise
Multipurpose_Motorized	How much trip purposes have motorized mode	Derived	N.A.	How much of the previous trip purposes have motorized mode, from 0 to 5
Multipurpose_PT	How much trip purposes have PT mode	Derived	N.A.	How much of the previous trip purposes have PT mode, from 0 to 5
Multipurpose_Unmotorized	How much trip purposes have unmotorized mode	Derived	N.A.	How much of the previous trip purposes have unmotorized mode, from 0 to 5

Novehicles_Motorized	User has no vehicles but uses motorized mode	Derived	N.A.	1 if user has no vehicles but is using at least one motorized mode for any trip purpose, 0 otherwise
Shopping_Missing	Missing mode for shopping trip purpose	Derived	N.A.	0 if missing value, 1 otherwise
Unimodal_Binary	The user is using the same mode for all trip purposes	Derived	N.A.	1 if the user is using the same mode for all trip purposes, 0 otherwise
Vehicles_Unmotorized	User has vehicles but uses unmotorized mode	Derived	N.A.	1 if user has vehicles but is using unmotorized mode for every trip purpose, 0 otherwise
Visiting_Missing	Missing mode for visting trip purpose	Derived	N.A.	0 if missing value, 1 otherwise
Attitudinal Indicators				
1_TITLE	On a scale of 1 to 5 where 1 strongly disagree and 5 strongly agree, indicate your level of agreement with the following statements:	Survey	1_TITLE	On a scale of 1 to 5 where 1 strongly disagree and 5 strongly agree, indicate your level of agreement with the following statements:
1_AUTONOMY	Here I can do all my activities autonomously	Survey	1_AUTONOMY	Here I can do all my activities autonomously
1_BELONGING	I feel like I belong to this area	Survey	1_BELONGING	I feel like I belong to this area
1_CAR	There are too many cars illegally parked on the sidewalks	Survey	1_CAR	There are too many cars illegally parked on the sidewalks
1_DISLIKE	I do not like spending much time in this area	Survey	1_DISLIKE	I do not like spending much time in this area
1_HELP	My neighbors are always willing to help each other	Survey	1_HELP	My neighbors are always willing to help each other
1_HOT	I avoid walking on the street on hot summer days	Survey	1_HOT	I avoid walking on the street on hot summer days
1_INFORMATION	I feel well informed about what is happening in the community (e.g., events and other activities)	Survey	1_INFORMATION	I feel well informed about what is happening in the community (e.g., events and other activities)
1_ISOLATION	I frequently feel isolated from others	Survey	1_ISOLATION	I frequently feel isolated from others
1_JOB	If I wanted to, I could find a job in this area	Survey	1_JOB	If I wanted to, I could find a job in this area
1_LEISURE	I regularly do my leisure activities in this area	Survey	1_LEISURE	I regularly do my leisure activities in this area
1_LIVEOTHERPLACE	I would not want to live anywhere else	Survey	1_LIVEOTHERPLACE	I would not want to live anywhere else
1_NEIGHBOR	I have a strong friendship with my neighbors	Survey	1_NEIGHBOR	I have a strong friendship with my neighbors
1_OLD	I can imagine myself aging comfortably in this area	Survey	1_OLD	I can imagine myself aging comfortably in this area
1_OTHERACTIVITIES	I usually do most of my other daily tasks in this area	Survey	1_OTHERACTIVITIES	I usually do most of my other daily tasks in this area
1_PARKS	I spend a lot of time outdoors in my area (e.g., walking, doing sports, in parks and gardens, etc.)	Survey	1_PARKS	I spend a lot of time outdoors in my area (e.g., walking, doing sports, in parks and gardens, etc.)
1_PRIDE	I feel proud to tell others that I live in this area	Survey	1_PRIDE	I feel proud to tell others that I live in this area
1_SHOP	I usually do my shopping in this area	Survey	1_SHOP	I usually do my shopping in this area
1_SLEEP	I practically only come here to sleep	Survey	1_SLEEP	I practically only come here to sleep
1_SLOPE	The slope of the streets makes it difficult to walk	Survey	1_SLOPE	The slope of the streets makes it difficult to walk

1_SOCIALLIFE	My social life is strongly centered in this area	Survey	1_SOCIALLIFE	My social life is strongly centered in this area
1_TRUST	People here are trustworthy	Survey	1_TRUST	People here are trustworthy
1_VOLUNTER	I usually participate in local community events and associative or volunteer activities	Survey	1_VOLUNTER	I usually participate in local community events and associative or volunteer activities
2_TITLE	Satisfaction with quantity/ease of walking access of the following amenities:	Survey	2_TITLE	Satisfaction with quantity/ease of walking access of the following amenities:
2_ASSOCI_ACESS	Satisfaction with cultural and sports community associations access	Survey	2_ASSOCI_ACESS	Satisfaction with cultural and sports community associations access
2_ASSOCI_QUANT	Satisfaction with cultural and sports community associations quantity	Survey	2_ASSOCI_QUANT	Satisfaction with cultural and sports community associations quantity
2_BATHROOM_ACESS	Satisfaction with public restrooms access	Survey	2_BATHROOM_ACESS	Satisfaction with public restrooms access
2_BATHROOM_QUANT	Satisfaction with public restrooms quantity	Survey	2_BATHROOM_QUANT	Satisfaction with public restrooms quantity
2_BUS_ACESS	Satisfaction with public transport stations and stops access	Survey	2_BUS_ACESS	Satisfaction with public transport stations and stops access
2_BUS_QUANT	Satisfaction with public transport stations and stops quantity	Survey	2_BUS_QUANT	Satisfaction with public transport stations and stops quantity
2_CAFE_ACESS	Satisfaction with cafes, restaurants and snack bars access	Survey	2_CAFE_ACESS	Satisfaction with cafes, restaurants and snack bars access
2_CAFE_QUANT	Satisfaction with cafes, restaurants and snack bars quantity	Survey	2_CAFE_QUANT	Satisfaction with cafes, restaurants and snack bars quantity
2_CENTER_ACESS	Satisfaction with community centers access	Survey	2_CENTER_ACESS	Satisfaction with community centers access
2_CENTER_QUANT	Satisfaction with community centers quantity	Survey	2_CENTER_QUANT	Satisfaction with community centers quantity
2_HEALTH_ACESS	Satisfaction with family health units, health centers and medical clinics access	Survey	2_HEALTH_ACESS	Satisfaction with family health units, health centers and medical clinics access
2_HEALTH_QUANT	Satisfaction with family health units, health centers and medical clinics quantity	Survey	2_HEALTH_QUANT	Satisfaction with family health units, health centers and medical clinics quantity
2_KINDER_ACESS	Satisfaction with kindergardens and primary schools access	Survey	2_KINDER_ACESS	Satisfaction with kindergardens and primary schools access
2_KINDER_QUANT	Satisfaction with kindergardens and primary schools quantity	Survey	2_KINDER_QUANT	Satisfaction with kindergardens and primary schools quantity
2_MARKET_ACESS	Satisfaction with grocery stores, mini-markets and supermarkets access	Survey	2_MARKET_ACESS	Satisfaction with grocery stores, mini-markets and supermarkets access
2_MARKET_QUANT	Satisfaction with grocery stores, mini-markets and supermarkets quantity	Survey	2_MARKET_QUANT	Satisfaction with grocery stores, mini-markets and supermarkets quantity
2_PARK_ACESS	Satisfaction with parks, green areas, gardens and viewpoints access	Survey	2_PARK_ACESS	Satisfaction with parks, green areas, gardens and viewpoints access
2_PARK_QUANT	Satisfaction with parks, green areas, gardens and viewpoints quantity	Survey	2_PARK_QUANT	Satisfaction with parks, green areas, gardens and viewpoints quantity
2_PHARMACIE_ACESS	Satisfaction with pharmacies access	Survey	2_PHARMACIE_ACESS	Satisfaction with pharmacies access
2_PHARMACIE_QUANT	Satisfaction with pharmacies quantity	Survey	2_PHARMACIE_QUANT	Satisfaction with pharmacies quantity

2_PLAYGROUND_ACESS	Satisfaction with playgrounds and leisure areas for children access	Survey	2_PLAYGROUND_ACESS	Satisfaction with playgrounds and leisure areas for children access
2_PLAYGROUND_QUANT	Satisfaction with playgrounds and leisure areas for children quantity	Survey	2_PLAYGROUND_QUANT	Satisfaction with playgrounds and leisure areas for children quantity
2_SQUARE_ACESS	Satisfaction with squares and plazas access	Survey	2_SQUARE_ACESS	Satisfaction with squares and plazas access
2_SQUARE_QUANT	Satisfaction with squares and plazas quantity	Survey	2_SQUARE_QUANT	Satisfaction with squares and plazas quantity
8_TITLE	On a scale of 1 to 5 where 1 strongly disagree and 5 strongly agree, indicate your level of agreement with the following statements:	Survey	8_TITLE	On a scale of 1 to 5 where 1 strongly disagree and 5 strongly agree, indicate your level of agreement with the following statements:
8_FALLINGDRY	I always feel safe walking on the street without falling when the pavement is dry	Survey	8_FALLINGDRY	I always feel safe walking on the street without falling when the pavement is dry
8_FALLINGWET	I always feel safe walking on the street without falling when the pavement is wet	Survey	8_FALLINGWET	I always feel safe walking on the street without falling when the pavement is wet
8_RUNOVER	I always feel safe from being run over when walking on the street	Survey	8_RUNOVER	I always feel safe from being run over when walking on the street
8_SAFEDAY	I always feel safe when I walk alone during the day	Survey	8_SAFEDAY	I always feel safe when I walk alone at night
8_SAFENIGHT	I always feel safe when I walk alone at night	Survey	8_SAFENIGHT	I always feel safe when I walk alone during the day
9_TITLE	How often do you visit/use the following amenities:	Survey	9_TITLE	How often do you visit/use the following amenities:
9_ASSOCI	Cultural and sports community associations	Survey	9_ASSOCI	Cultural and sports community associations
9_BATHROOM	Public restrooms	Survey	9_BATHROOM	Public restrooms
9_BUS	Public transport stations and stops	Survey	9_BUS	Public transport stations and stops
9_CAFE	Cafes, Restaurants or snack bars	Survey	9_CAFE	Cafes, Restaurants or snack bars
9_CENTER	Day Centers/Community Centers	Survey	9_CENTER	Day Centers/Community Centers
9_HEALTH	Family Health Units, Health Centers, Medical Clinics	Survey	9_HEALTH	Family Health Units, Health Centers, Medical Clinics
9_KINDER	Kindergartens/Primary Schools	Survey	9_KINDER	Kindergartens/Primary Schools
9_MARKET	Grocery stores, Mini-markets or Supermarkets	Survey	9_MARKET	Grocery stores, Mini-markets or Supermarkets
9_PARK	Parks, Green areas, Gardens or Viewpoints	Survey	9_PARK	Parks, Green areas, Gardens or Viewpoints
9_PHARMACIE	Pharmacies	Survey	9_PHARMACIE	Pharmacies
9_PLAYGROUND	Playgrounds or leisure areas for children	Survey	9_PLAYGROUND	Playgrounds or leisure areas for children
9_SQUARE	Squares, Plazas or Small Squares (for events, social gatherings and urban life)	Survey	9_SQUARE	Squares, Plazas or Small Squares (for events, social gatherings and urban life)
Access_15min	Average of some ease of access indicators (2)	Derived	N.A.	Average of "2" questions
Access_15min_Ordinal	Ordinal value for previous variable	Derived	N.A.	Dividing Access_15min in 5 categories, from 1 to 5
Access_15min_Ordinal_Opt	Recoding of the previous variable into fewer categories	Derived	N.A.	Dividing Access_15min in 3 categories, from 0 to 2
Access_15min_Ordinal_Opt2	Recoding of the previous variable into fewer categories	Derived	N.A.	Dividing Access_15min in 2 categories, from 0 to 1

Access_New	Average of the ease of access indicators that were more significant after factor analysis (1)	Derived	N.A.	Average of selected "2" questions
Access_New_Ordinal	Ordinal value for previous variable	Derived	N.A.	Dividing Access_New in 5 categories, from 1 to 5
Air	Air satisfaction	Survey	5_AIR	Considering a scale of 1 to 5 where 1 means you are very dissatisfied and 5 – very satisfied, how satisfied are you regarding air quality
Bench	Bench availability	Survey	3_BENCH	Considering a scale of 1 to 5 where 1 means you are very dissatisfied and 5 – very satisfied, how satisfied are you with the quantity of benches or seating areas
Building	Appearance of buildings	Survey	4_BUILDINGS	Considering a scale of 1 to 5 where 1 means you are very dissatisfied and 5 – very satisfied, how satisfied are you regarding the general exterior appearance of the buildings (e.g., appearance of facades, cracks, integrity of windows, signs of abandonment).
Clean	Cleanliness satisfaction	Survey	5_CLEAN	Considering a scale of 1 to 5 where 1 means you are very dissatisfied and 5 – very satisfied, how satisfied are you regarding street cleanliness
Clean_Opt	Recoding of the previous variable into fewer categories	Derived	N.A.	0 if <4, 1 otherwise
Density	Density satisfaction	Survey	4_DENSITY	Considering a scale of 1 to 5 where 1 means you are very dissatisfied and 5 – very satisfied, how satisfied are you regarding the relationship between building density and open spaces (is the area very dense or are there a good number of squares and green spaces, wide streets, etc.)
IND_Access_15min	Recoding of the previous variable into fewer categories	Derived	N.A.	Dividing Access_New in 3 categories, from 0 to 2
IND_Living_15min	Recoding of the previous variable into fewer categories	Derived	N.A.	Dividing Living_New in 3 categories, from 0 to 2
IND_Safety_15min	Recoding of the previous variable into fewer categories	Derived	N.A.	Dividing Living_15min in 3 categories, from 0 to 2
Light	Lighting availability	Survey	3_LIGHT	Considering a scale of 1 to 5 where 1 means you are very dissatisfied and 5 – very satisfied, how satisfied are you with the quantity of public lighting
Living_15min	Average of some neighbourhood living satisfaction indicators (1)	Derived	N.A.	Average of "1" questions
Living_15min_Ordinal	Ordinal value for previous variable	Derived	N.A.	Dividing Living_15min in 5 categories, from 1 to 5
Living_15min_Ordinal_Opt2	Recoding of the previous variable into fewer categories	Derived	N.A.	Dividing Living_15min in 2 categories, from 0 to 1
Living_New	Average of the neighbourhood living satisfaction indicators that were more significant after factor analysis (1)	Derived	N.A.	Average of selected "1" questions
Living_New_Ordinal	Ordinal value for previous variable	Derived	N.A.	Dividing Living_New in 5 categories, from 1 to 5
Maintenance	Maintenance of green areas	Survey	4_MANTAIN	Considering a scale of 1 to 5 where 1 means you are very dissatisfied and 5 – very satisfied, how satisfied are you regarding the maintenance of parks, green areas, gardens, viewpoints

Noise	Noise satisfaction	Survey	5_NOISE	Considering a scale of 1 to 5 where 1 means you are very dissatisfied and 5 – very satisfied, how satisfied are you regarding noise level
Parking	Parking satisfaction	Survey	4_PARKING	Considering a scale of 1 to 5 where 1 means you are very dissatisfied and 5 – very satisfied, how satisfied are you regarding the way people park, in general (e.g., do they frequently obstruct sidewalks or traffic flow?)
People	Number of people on the street	Survey	5_PEOPLE	Considering a scale of 1 to 5 where 1 means you are very dissatisfied and 5 – very satisfied, how satisfied are you regarding number of people on the street
Safety_15min	Average of some safety indicators (8)	Derived	N.A.	Average of "8" questions
Safety_15min_Ordinal	Ordinal value for previous variable	Derived	N.A.	Dividing Safety_15min in 5 categories, from 1 to 5
Safety_15min_Ordinal_Opt	Recoding of the previous variable into fewer categories	Derived	N.A.	Dividing Safety_15min in 3 categories, from 0 to 2
Safety_15min_Ordinal_Opt2	Recoding of the previous variable into fewer categories	Derived	N.A.	Dividing Safety_15min in 2 categories, from 0 to 1
Satisfaction	Overall satisfaction	Survey	7_GLOBAL	Considering a scale of 1 to 5 where 1 means you are very dissatisfied and 5 – very satisfied, how satisfied are you globally with the area surrounding your residence?
Sidewalk	Sidewalk satisfaction	Survey	4_SIDEWALK	Considering a scale of 1 to 5 where 1 means you are very dissatisfied and 5 – very satisfied, how satisfied are you regarding the quality and general condition of the sidewalks (e.g., deterioration, obstructions, and pavement material)
Traffic	Traffic satisfaction	Survey	5_TRAFFIC	Considering a scale of 1 to 5 where 1 means you are very dissatisfied and 5 – very satisfied, how satisfied are you regarding intensity of car traffic
Traffic_Opt	Recoding of the "Traffic" variable into fewer categories	Derived	N.A.	0 if <4, 1 otherwise
Tree	Tree availability	Survey	3_TREE	Considering a scale of 1 to 5 where 1 means you are very dissatisfied and 5 – very satisfied, how satisfied are you with the quantity of trees on streets and in other public spaces
Use_15min	Average of some amenity use indicators (9)	Derived	N.A.	Average of "9" questions
Use_15min_Ordinal	Ordinal value for previous variable	Derived	N.A.	Dividing Use_15min in 5 categories, from 1 to 5
Use_15min_Ordinal_Opt	Recoding of the previous variable into fewer categories	Derived	N.A.	Dividing Use_15min in 3 categories, from 0 to 2
Use_15min_Ordinal_Opt2	Recoding of the previous variable into fewer categories	Derived	N.A.	Dividing Use_15min in 2 categories, from 0 to 1

A2: Python Code

The following annex illustrates the Python script employed to estimate the Integrated Choice and Latent Variable (ICLV) models. Using the Biogeme package, the code defines the structural equations for both the unobserved psychological constructs and the modal utilities. These are simulated simultaneously via the Monte Carlo integration using 1000 Halton draws. Furthermore, it includes the measurement equations for the survey indicators that shape the latent variables. The script runs an automated Maximum Likelihood Estimation (MLE) across all investigated trip purposes (commuting, shopping, visiting and 15-min) and exports the goodness of fit indicators, the parameter coefficients and their significance.

```
import pandas as pd
import biogeme.database as db
import biogeme.biogeme as bio
from biogeme import models
from biogeme.expressions import Beta, Variable, Draws, MonteCarlo, log, exp
from biogeme.results_processing import get_pandas_estimated_parameters
import math
import os
import glob

N_DRAWS = 1000
N_THREADS = 6

def run_model(csv_path, choice_col, output_name, model_name):
    print(f"\nEstimating {model_name}...")

    # Load data
    try:
        df = pd.read_csv(csv_path, sep=None, engine='python', encoding='latin1')
    except:
        df = pd.read_csv(csv_path, sep=";", encoding='latin1')

    # Map choice variable
    choice_map = {"Motorized": 1, "Unmotorized": 2, "Public Transport": 3}
    df["CHOICE"] = df[choice_col].astype(str).str.strip().map(choice_map)
```

```python
df.dropna(subset=["CHOICE"], inplace=True)
df["CHOICE"] = df["CHOICE"].astype(int)

# Factor Scores
indicators = ["Factor_Living_15min", "Factor_Access_15min", "Factor_Safety_15min"]

# Select columns
cols = (["CHOICE", "Age", "Age_sq", "Education_Level_Ordinal_Opt",
        "Clean_Opt", "Dummy_Povoa", "Dummy_Sobr", "MV_Ordinal",
        "PT_Card_Binary", "Driver_License_Binary", "Traffic", "Parking"]
        + indicators)

# Create database
database = db.Database(output_name.replace(".csv", "_db"), df[cols])

# Define variables
CHOICE = Variable("CHOICE")
Age = Variable("Age")
Age_sq = Variable("Age_sq")
Education_Level_Ordinal_Opt = Variable("Education_Level_Ordinal_Opt")
PT_Card_Binary = Variable("PT_Card_Binary")
Driver_License_Binary = Variable("Driver_License_Binary")
Clean_Opt = Variable("Clean_Opt")
Dummy_Povoa = Variable("Dummy_Povoa")
Dummy_Sobr = Variable("Dummy_Sobr")
MV_Ordinal = Variable("MV_Ordinal")

Factor_Living_15min = Variable(indicators[0])
Factor_Access_15min = Variable(indicators[1])
Factor_Safety_15min = Variable(indicators[2])
Traffic = Variable("Traffic")
Parking = Variable("Parking")


# Latent variables and measurement equations
omega1 = Draws("omega1", "NORMAL_HALTON2")
omega2 = Draws("omega2", "NORMAL_HALTON3")

LV1 = (
```

```
            Beta("B_Age_LV1", 0, None, None, 0) * Age +
            Beta("S_LV1", 1, None, None, 1) * omega1)

    LV2 = (
            Beta("S_LV2", 1, None, None, 1) * omega2)

    # Gauss function
    def gauss_like(y, name, lv):
        beta = Beta(f"b_{name}", 1, None, None, 0)
        alpha = Beta(f"a_{name}", 0, None, None, 0)
        sigma = Beta(f"s_{name}", 1, None, None, 0)
        mu = alpha + beta * lv
        return (1.0 / math.sqrt(2 * math.pi) / sigma) * exp(-(y - mu) ** 2 / (2 * sigma ** 2))

    # Measurement equations for LV1
    like_lv1 = gauss_like(Factor_Living_15min, "living", LV1) * gauss_like(Factor_Access_15min, "access", LV1) *
gauss_like(Factor_Safety_15min, "safety", LV1)
    like_lv2 = gauss_like(Traffic, "traffic", LV2) * gauss_like(Parking, "parking", LV2)

    # Utility parameters
    ASC_M = Beta("ASC_M", 0, None, None, 1)
    ASC_U = Beta("ASC_U", 0, None, None, 0)
    ASC_PT = Beta("ASC_PT", 0, None, None, 0)

    V_M = (
            ASC_M +
            Beta("B_Education_Level_Ordinal_Opt_M", 0, None, None, 0) * Education_Level_Ordinal_Opt +
            Beta("B_MV_Ordinal_1_M", 0, None, None, 0) * (MV_Ordinal == 1) +
            Beta("B_MV_Ordinal_2+_M", 0, None, None, 0) * (MV_Ordinal >= 2) +
            Beta("B_Driver_License_Binary_M", 0, None, None, 0) * Driver_License_Binary)

    V_U = (
            ASC_U +
            Beta("L_LV1_U", 0, None, None, 0) * LV1 +
            Beta("L_LV2_U", 0, None, None, 0) * LV2 +
            Beta("B_Age_U", 0, None, None, 0) * Age +
            Beta("B_Age_sq_U", 0, None, None, 0) * Age_sq +
            Beta("B_Dummy_Povoa_U", 0, None, None, 0) * Dummy_Povoa)
```

```
    V_PT = (
            ASC_PT +
            Beta("B_Age_PT", 0, None, None, 0) * Age +
            Beta("B_Age_sq_PT", 0, None, None, 0) * Age_sq +
            Beta("B_PT_Card_Binary_PT", 0, None, None, 0) * PT_Card_Binary +
            Beta("B_Clean_Opt_PT", 0, None, None, 0) * Clean_Opt +
            Beta("B_Dummy_Sobr_PT", 0, None, None, 0) * Dummy_Sobr)

    V = {1: V_M, 2: V_U, 3: V_PT}

    # Estimation
    loglike = log(MonteCarlo(models.logit(V, {1: 1, 2: 1, 3: 1}, CHOICE) * like_lv1 * like_lv2))
    bgm = bio.BIOGEME(database, loglike, number_of_draws=N_DRAWS)
    bgm.model_name = model_name
    bgm.numberOfThreads = N_THREADS
    res = bgm.estimate()

    if res:
        table = get_pandas_estimated_parameters(res)
        table.rename(columns={"Value": "Coefficient", "Robust std err.": "Std. Err.",
                              "Robust t-stat.": "t-stat", "Robust p-value": "p-value"}, inplace=True)
        table.to_csv(output_name, index=True, sep=';', decimal='.')
        print(f"Saved: {output_name}")
        return res.get_general_statistics()

if __name__ == "__main__":
    tasks = [
        ("Data_Commuting.csv", "Mode_commuting_Opt", "ICLV_Commuting_Results.csv", "ICLV_Commuting"),
        ("Data_Shopping.csv", "Mode_shopping_Opt", "ICLV_Shopping_Results.csv", "ICLV_Shopping"),
        ("Data_Visiting.csv", "Mode_visiting_Opt", "ICLV_Visiting_Results.csv", "ICLV_Visiting"),
        ("Data_15min.csv", "Mode_15mincity_Opt", "ICLV_15min_Results.csv", "ICLV_15min")
    ]
    gof_results = {}
    # Run all models
    for data_file, choice_col, out_file, mod_name in tasks:
        stats = run_model(data_file, choice_col, out_file, mod_name)
        if stats is not None:
            mod_short = mod_name.split("_")[1]
            gof_results[mod_short] = {
```

```
                "Sample size": stats.get('Sample size', None),
                "Number of estimated parameters": stats.get('Number of estimated parameters', None),
                "Init log likelihood": stats.get('Init log likelihood', None),
                "Final log likelihood": stats.get('Final log likelihood', None),
                "Rho-square for the init. model": stats.get('Rho-square for the init. model', None),
                "Rho-square-bar for the init. model": stats.get('Rho-square-bar for the init. model', None),
                "Akaike Information Criterion": stats.get('Akaike Information Criterion', None),
                "Bayesian Information Criterion": stats.get('Bayesian Information Criterion', None)
            }

# Create summary files
print("\nCreating summary files...")
summary_dfs = []
pub_dfs = []
for _, _, out_file, mod_name in tasks:
    if os.path.exists(out_file):
        df_res = pd.read_csv(out_file, sep=";", decimal=".")
        df_res.set_index("Name", inplace=True)
        mod_short = mod_name.split("_")[1]

        df_full = df_res[["Coefficient", "Std. Err.", "t-stat", "p-value"]].copy()
        df_full.columns = [f"{mod_short} {c}" for c in df_full.columns]
        summary_dfs.append(df_full)

        def format_pval(row):
            v, p = row['Coefficient'], row['p-value']
            s = f"{v:.3f}"
            if pd.isna(p):
                return s
            elif p < 0.01:
                return s + "***"
            elif p < 0.05:
                return s + "**"
            elif p < 0.10:
                return s + "*"
            return s

        df_pub = pd.DataFrame(index=df_res.index)
        df_pub[mod_short] = df_res.apply(format_pval, axis=1)
```

```
            pub_dfs.append(df_pub)

    if summary_dfs:
        pd.concat(summary_dfs, axis=1).to_csv("ICLV_All_Results_Summary.csv", sep=";", decimal=".", index=True)
        pd.concat(pub_dfs, axis=1).to_csv("ICLV_Significance_Table.csv", sep=";", decimal=".", index=True)
    if gof_results:
        pd.DataFrame(gof_results).T.map(lambda x: x[0] if isinstance(x, tuple) else
x).to_csv("ICLV_Goodness_Of_Fit_Summary.csv", sep=";", decimal=".", index=True)

    # Clean
    protected = [t[2] for t in tasks] + ["ICLV_All_Results_Summary.csv", "ICLV_Publication_Table.csv",
"ICLV_Goodness_Of_Fit_Summary.csv"]
    for ext in ["*.html", "*.yaml", "*.pickle", "biogeme.toml", "*.iter"]:
        for f in glob.glob(ext):
            if f not in protected:
                try: os.remove(f)
                except: pass

    print("\nDone")
```