Attitudes and Latent Class Choice Models using Machine Learning: An application in Car-Sharing

Lorena Torres Lahoz - s202392

Supervisors:
Carlos M.Lima Azevedo
Mayara Monteiro
Georges Sfeir

January, 2022

Attitudes and Latent Class Choice Models using Machine Learning: An application in Car-Sharing

Master thesis
January, 2022

By Lorena Torres Lahoz

ECTS:	30
Project period:	August 2021-January 2022
Department:	Technology, Management and Economics Transport
Supervisors:	Carlos M.Lima Azevedo
	Mayara Monteiro
	Georges Sfeir

Approval

This thesis has been prepared over six months at the Department of Department of Technology, Management and Economics Transport, at the Technical University of Denmark, DTU, in partial fulfilment for the degree Master of Science in Engineering, MSc Eng.

Lorena Torres Lahoz - s202392

....

Signature

....

Date

Acknowledgements

I would like to thank all the help and motivation received from Carlos, Mayara, George and Ioanna, that guided me along this exciting way.
This work would not have been possible without the unconditional support of my parents, to whom I will always be grateful.
Thanks also to my friends, who have been there for me no matter the distance, and thanks to those who have become my family during my time in Denmark.
Finally, thanks to those who are no longer with us but who had believed in me since the beginning.

Abstract

Latent Class Choice Models (LCCM) are extensions of general discrete choice models (DCMs) typically implemented to capture unobserved heterogeneity in the choice process by segmenting the population based on similarities. Socio-economic characteristics of the decision-makers have typically defined latent classes. However, incorporating individuals' attitudes or beliefs, measured using psychometric indicators, into the specification of the latent classes can offer additional behavioral insights, allowing for more realistic market segmentation and more specific design policies. This study analyzes different methods of incorporating attitudinal indicators in the discrete choice model employing machine learning (ML) techniques, namely Embeddings and Gaussian-Bernoulli Mixtures. Without compromising the economic and behavioral interpretability of the models, ML-based techniques offer us a powerful way to capture unobserved behavioral patterns. Models with and without attitudinal variables are compared to evidence the relevance of attitudes in the estimation and class configuration of LCCMs.

The application aims to model people's behavior towards different types of car sharing. By including travelers' attitudes into the models, we can implement personalized policies to attract and retain car-sharing members while aiding in its sustainable integration with the existing transportation system.

Keywords

Machine learning, Latent Class Choice Models, Discrete Choice Models, car-sharing, psychometric indicators, Gaussian-Bernoulli Mixture Latent Class Choice Model, Embeddings.

Contents

1 Introduction

Our project aims to explore and investigate the inclusion of attitudinal statements in latent class choice models. We analyze different model formulations with and without the inclusion of attitudinal variables to understand better how believes and attitudes influence the distribution of the latent classes in discrete choice models.
For our project application, we aim to model people's behavior towards different types of car sharing. By analyzing the travelers' preferences and attitudes, we can provide personalized incentives and service features for different population segments that attract and retain car-sharing members while aiding car-sharing sustainable integration with the existing transportation system [1]
There are many examples in the transportation literature where the segmentation of the population into latent groups allows for a convenient, flexible, and intuitive way to introduce taste heterogeneity in choice models [2]. Latent classes refer to unobservable population groups, where each individual has an associated probability of belonging to each group/class. This segmentation leads to more accurate and applicable results, such as more efficient design policies.
The traditional latent class choice models based their segmentation on socio-demographic variables. Some previous studies have focused, for example, on how socio-demographic variables influence the probability of becoming a car-sharing member [10][11]. However, this approach is minimal.
The introduction into the models of attitudinal variables that reflect the attitudes of individuals on subjects related to the choice allows for a more complete and richer analysis of the latent classes. Our analysis evidence that it also provides different latent class configurations and estimates, staying that attitudinal information is significant for the decision process. It is the case when within a group that shares similar socio-demographic variables, differences in their behavior towards car-sharing still exist. Some other studies have also proved that the estimation of latent class models can be improved using psychometric indicators that measure the effect of unobserved attitudes [2].
Our application is in the field of car-sharing, which offers short-term car access to its users, providing a flexible alternative that meets diverse transportation needs while reducing the negative impacts of private vehicle ownership. For nearly 20 years, car-sharing usage has been growing. In 2018, car-sharing was operating in 47 countries and six continents, with over 98.000 vehicles available for more than 32 million members [3]
This new business idea enables its users to access a car for a certain amount of time and endows them greater mobility flexibility. It provides their users a reduction in their commutes times and increases their use of other alternative transportation modes [5]. Moreover, this availability of cars provokes that car-sharing members own fewer vehicles than the rest of the population [4]. They also affect positively both socially and environmentally in cities given its capacity to reduce parking demand [9], and congestion [8], which results in a decrease of greenhouse gases and air pollutant emissions by reducing the kilometers traveled [6]. Nevertheless, car-sharing personal monetary savings has primarily contributed to its growth over the past decade [5].

We analyze individual behavior through different car-sharing businesses, depending on their possibilities of pick-up and return locations, such as station-based, free-floating, or round-trip. Besides, other attributes can change depending on the car-sharing company as the type of car engine, the pricing scheme, the ownership of the cars, or the parking availability.
Our study is based on data collected in Copenhagen, a bicycle-friendly city. We employ data from a stated preference (SP) choice experiment, socio-demographic variables, and attitudinal questions toward car and car-sharing services. Moreover, it has been analyzed that local elements also affect car-sharing preferences, such as the use of bikes, the price of public transport, or the availability of car-sharing services. Service features ideal for one context may not apply to other cities. [12]. Thus, these local features are also incorporated in our models.
In conclusion, we explore different formulations to include attitudinal statements in discrete choice models. We evidence improvements in the estimation and configuration of the classes when attitudinal indicators are present in the models. We open the door for future investigation in models formulation that integrates latent classes and attitudinal variables.
For this concrete application, we dig into the impact of attitudinal questions on people's behavior towards car ownership and different types of car-sharing businesses. We contribute to the car-sharing literature by detecting the difference in car-sharing preferences when targeted to different population segments sharing similar beliefs and attitudes. This segmentation allows us to design more tailor-made policies to subscribe and retain satisfied car-sharing members, optimizing car-sharing business integration.

2 Literature review

Understanding how people make choices is a beneficial information for policymaking, designing marketing campaigns, planning the transportation system in cities, and many other situations. Therefore, much effort has been made to understand the behavioral process that leads to the choice made by an agent.
The causal perspective seems the most convenient, where some that factors collectively determine or cause the agent's choice. Some of these factors are observed by the researcher, and some are not. Thus, the agent's choice can model through a function $y = h(x, \epsilon)$, called behavioral process, where ϵ is the not observed part of the decision. Therefore, the agent's choice is not deterministic and cannot be precisely predicted. Instead, we get the probability of choosing each alternative. [23]
Thurstone (1927) first developed these concepts in terms of psychological stimuli, leading to a binary probit model [25] and Marschak (1960) interpreted stimuli as utility and provided a formulation based on utility maximization, called random utility models (RUM) [26]. Although, the theory of discrete choice was finally consolidated with the contribution of McFadden, work that earned him the Nobel Prize in Economics in 2000. He proposed a multinomial logit model in connection with the consumer demand theory.
The main principle of RUM is that the individual has a utility function Ui associated with each alternative. He chooses the option that provides him with the maximum utility. The formulation of the utility of an individual n (U_{in}), includes factors connected with the alternative j, named as the attributes (X_{nj}), as well as socio-characteristics variables of the decision-maker (S_n). This observed part of the utility is called the representative utility, V_{nj}. However, not all the variables can be observed by the researcher when a person makes a decision. For this reason, a disturbance term is included in the formulation of the utility, taking into account these unobserved factors. The disturbance term, called ϵ, is not defined for a choice situation per se, and the researcher can modify its specification.

$$U_{nj} = V_{nj} + \epsilon_{nj} \tag{1}$$

It is worth mentioning that only the differences between utilities (U_{nj}) matter. Therefore, alternative-specific coefficients can be included to measure the average effect of no included factors in the utility of an alternative compared with all other ones.

2.1 Logit model

In the logit model, the error term ϵ is assumed to be independent and identically distributed following an extreme value distribution, in this way, the final model is computationally solvable with an interpretable and closed expression. The density for the unobserved part of each utility is, therefore, defined as:

$$f(\epsilon_{nj}) = e^{-\epsilon_{nj}} e^{-e^{-\epsilon_{nj}}} \tag{2}$$

and its cumulative distribution is

$$F(\epsilon_{nj}) = e^{-e^{-\epsilon_{nj}}} \tag{3}$$

The difference between two variables of type extreme value, $\epsilon^*_{nji} = \epsilon_{nj} - \epsilon_{ni}$ follows a logarithmic distribution,

$$F(\epsilon^*_{nji}) = \frac{e^{\epsilon^*_{nji}}}{1 + e^{\epsilon^*_{nji}}} \tag{4}$$

Thus, the probability that the decision maker n chooses alternative i can be expressed as:

$$P_{ni} = Prob(V_{vi} + \epsilon_{ni} > V_{nj} + \epsilon_{nj} \forall j \neq i) = Prob(\epsilon_{nj} < \epsilon_{ni} + V_{ni} - V_{nj} \forall j \neq i) \tag{5}$$

that after some manipulations turns out in the closed expression,

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}} \tag{6}$$

called the logit probability.

Later on, the development of discrete choices models has led to different models, such as generalized extreme value models, probit, and mixed logit, valid for modeling in different situations. In addition, they also overcome some of the restrictions present in logit models, such as independence of irrelevant alternatives (IIA) or random variation of preferences.
IIA means that the ratio between alternatives i and k, only depends on these alternatives, and it remains constant, no matter what other alternatives appear.

$$\frac{P_{ni}}{P_{nk}} = \frac{\frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}}}{\frac{e^{V_{nk}}}{\sum_j e^{V_{nj}}}} = \frac{e^{V_{ni}}}{e^{V_{nk}}} = e^{V_{ni} - V_{nk}} \tag{7}$$

In other words, an improvement in one alternative is proportionally extracted from all remaining alternatives. For example, if the probability of one choice improves by 20%, the probabilities of all remaining alternatives fall by 20%. This pattern may be an unrealistic and restrictive assumption to model in real situations.
Moreover, the variation in the preferences between individuals can only be captured if they vary systematically given the observed variables. For a random change of taste, logit is not a good approximation of the reality since the unobserved part of the utility is assumed to be independent and, therefore, does not provide information on the error of a different alternative.
Finally, logit is not a good model specification when dealing with panel data, i.e. when multiple choices are made by the same individual. Because just as some dynamics are expected in the observed variables, there are also dynamics in the unobserved ones and it is hard to assume that they are independent [24].
In conclusion, the logit model is the ancestor of more complex models described hereafter.

2.2 Mixed logit

Our first model is a mixed logit. It is a highly flexible model that can approximate any random utility model [13]. This model has none of the limitations mentioned above of the logit model.
Mixed logit is any model in which probabilities can be expressed as,

$$P_{ni} = \int \left(\frac{e^{\beta' x_{ni}}}{\sum_{j=1}^{J} e^{\beta' x_{nj}}} \right) f(\beta) d\beta \tag{8}$$

where $f(\beta)$ is the probability density function of β. In other words, the mixed logit model probability is a weighted average of the logit formula evaluated at different beta values, where the weights are given by $f(\beta)$.
When the distribution of the beta values is discrete and takes a finite number of values, the mixed model is called latent class choice model. We will explain this model in the following section.
When the distribution of beta $f(\beta)$ is continuous, the beta coefficients change between decision-makers with density $f(\beta)$. The researcher specifies its distribution $f(\beta|\theta)$, according to its expectations of behavior, and estimates its parameters θ. Usually, the distribution is specified as normal or log-normal. This last one is useful if the coefficient sign is known; for example, the cost coefficient is always negative.
The increase in computation speed and the discovery of new estimation methods pushed the usage of mixed logit models.

2.3 Latent class choice models (LCCM)

It was Walker and Ben-Aliva (2002) [14], who presented a practical generalized random utility model with extensions for latent variables and latent classes. Extensions added to the basic RUM to relax assumptions and enrich the capabilities of the model.
Latent class choice models (LCCM) are employed when latent classes are discrete constructs. The hypothesis is that there may be discrete segments of decision-makers that are not immediately identifiable from the data. However, the population can be probabilistically segmented into groups that have different preferences, different decision protocols, or behave differently towards the decision, signified by class-specific utility equations for each class. [14]
Evidences in the literature suggests that latent class models are a very convenient, flexible, and intuitive way to introduce taste heterogeneity in discrete choice models. One of the first applications of LCCM was in 1995 when Gopinath modeled shippers' choices allowing for different sensitivities to time and cost [15].
The critical issue in latent class choice models is how to specify the class membership model since it is not defined forehand. Typically, straightforward logit equations are employed. We will explore different class membership models in the following sections.

2.4 Machine learning techniques

In recent years, machine learning techniques have been applied to choice models problems, resulting in increased prediction accuracy.
Nevertheless, machine learning techniques have often been criticized in economics and transportation for their lack of interpretability. While that may have been true in its early days, nowadays, many different studies have focused on improving prediction estimates without sacrificing the model's interpretability.
The choice models employing machine learning techniques that we will present in sections 3.4 and 3.6 are examples of this tendency.
We describe hereafter the two machine learning techniques implemented in our models.

2.4.1 Clustering

Clustering techniques belong to unsupervised learning, also called exploratory analysis, characterized by no output variables. Unsupervised learning aims to identify patterns in the data by learning some structure, finding regularities or discovering hidden information.
Clustering a set of observations means dividing them into non-overlapping groups. These groups are meaningful because they capture some data structure. Observations are grouped using similarity measures, and clusters interpretation is helpful for a specific purpose.
Different machine learning techniques can be applied for clustering. Following the formulation of [31], which we will explain in section 3.4, we chose a model-based clustering technique based on parametric mixture models for the formulation of the class membership model. Mixture models can represent the presence of subpopulations within an overall population. In mixture models, each data observation has an associated probability of belonging to each distribution, where each distribution represents a latent (unobserved) cluster/class.
The justification for employing mixture models is triple. Firstly, a probabilistic model is needed to estimate the latent classes simultaneously with the discrete choice model, which improves the estimation compared to sequential estimation. Secondly, mixture models provide more flexibility than the utility specification of latent classes commonly defined as linear functions. Thirdly, this clustering technique also allows for interpretable results when evaluating the clusters. [32].

2.4.2 Embeddings

For our last model formulation, we will employ continuous vector representation, called embeddings, to encode categorical or discrete explanatory variables with a particular focus on interpretability and model transparency.
Within the framework of DCMs, categorical variables are typically encoded using dummy encoding, also called 'one-hot-encoding.' However, this type of encoding, although computationally simple, has several disadvantages. Firstly, it increases the model's dimensionality proportionally to the cardinality of the categorical variables considered. A high-dimensional model is more exposed to suffer what is known as "the curse of dimensionality," which accounts for the challenge encountered when an increase in the number of dimensions requires an exponential increase in the sample size to maintain the model reliability. The amount of

data that contributes to estimating each coefficient becomes insufficient and worsens when the heterogeneity of the data is high.
At the same time, travel data collection typically has a small sample size. This fact, coupled with the increase of model dimensions, may lead to overfitting. Overfitting occurs when the model is so complex that it describes noise or random errors in the data instead of only the relation between the dependent and independent variables. Although an overfitted model presents a good estimation in the train set, it does not perform well on out-of-sample predictions where these noise terms do not follow a systematic pattern. Thus, if we do not consider overfitting when selecting a model, we risk choosing complex and non-generalized models.
On the other hand, the second disadvantage of dummy encoding is that it assumes that categorical variables are mutually exclusive and unrelated. It does not allow to model similarity representation between categories, which would be intuitive for a human to recognize. In other words, it ignores the contextual information behind the features.
Therefore, embeddings enable us to encode categorical variables to reduce the problems mentioned above by employing a data-driven method.
Embedding first became popular with the creation of word2vec, a deep learning-based method for generating a vector of words, called words embeddings. It became a valuable technique for capturing the semantic relatedness of words based on their distributional properties. Similar words have similar vectors in the embeddings space, allowing the formation of meaningful word clusters. Their first applications were in Natural Language Processing (NLP) which permits machines to break down and interpret human language. [30]
Embedding representations within the logit framework have been first conceptualized by Pereira in [28], where he mapped discrete variables used in travel demand modeling into a latent embedding space exceeding traditional methods as PCA or dummy encoding. However, we followed the formulation of Arkoudi in [27] that allow us to preserve the interpretability of the embeddings vectors by linking each of their dimensions to an alternative within the choice set. This formulation allows us to visualize and analyze the categorical variables in a meaningful continuous space.

2.5 Attitudinal variables in DCMs

Finally, we explore some of the literature connecting attitudinal indicators and latent class discrete choice models. We found examples in the literature on the relationship between attitudes and behavior in mode choices [16].
The inclusion of attitudinal variables in LCCM helps construct more realistic population segments that can help design more appropriate strategies and effective policies [17] [18]. Research into attitude-based segmentation has significantly increased in recent years since it is a successful area with potential improvements in the explanatory power of choice models. Psychometric indicators are additional information that can specify and estimate latent classes. They measure the effect of unobserved attributes in individuals' preferences on topics related to the choice. An example of a psychometric indicator can be: 'We need to build more parking lots downtown.' A positive answer to this statement shows a desire for a more car-based city, and a negative one a preference for a more environmentally friendly city.

In [20], a LCCM is estimated, and psychometric indicators are included in the maximum likelihood estimation to reinforce the model. The psychometric indicators are conditional on the latent class s. Therefore, the item-response probability of observing indicator I_n is given by $P_n(I_n|s)$ that is defined as a parameter jointly estimated with the choice and the class membership model. The model evidence that including the psychometric indicators allows for a richer analysis and generates significantly different estimates for the class membership model.
On the other hand, in [2], psychometric indicators are introduced by assuming that the probability of giving an agreement level I_K to the k_{th} attitudinal question also depends on the respondent's class s, $P_n(I_k|s)$. In this case, they use an ordinal logit approach to model the response probability $P_n(I_k|s)$, since the responses to the indicators consist of a few ordered integer values corresponding with the level of agreement with the statement on a Likert scale [19]. The advantage of this formulation is the close form of the ordinal logit used to measure the indicators that allow for a more straightforward estimation procedure, where the choice and the response to the indicators are estimated together.
Another study has gone further, studying the relationship between normative belief, modality styles, and travel behavior. Normative belief refers to the individual's perception of the opinion of others concerning a specific behavior. At the same time, modality style describes the part of an individual's lifestyle characterized by using a particular set of travel modes.
They show evidence of associations between travel behavior and different latent psychosocial constructs. In this case, latent normative belief can predict the engagement to different modality styles. Modality styles are represented as latent classes defined by mode-use frequencies, mobility attitudes, mode-specific attitudes, and habits. This model exemplifies how the socio-psychological approach, which originated in social psychology, can extend conventional travel behavior analysis. [21]
Finally in [22], exploratory factor analysis and latent class cluster analysis of attitudinal variables is performed to cluster individuals in groups concerning their inclination to adopt MaaS (Mobility as a Service). Even though the model is enriched with a series of covariates referring to socioeconomic, mobility, and technology-related characteristics, they do not improve the model; they only help cluster identification. Attitudinal variables are the ones defining the structure.
The previous examples have highlighted the importance and relevance of including attitudes, preferences, and beliefs in discrete choice models, especially in LCCM. They improve the model estimation and create more realistic and valuable models.

3 Model framework and formulation

This section will present the formulation and theory for all models that we have implemented, in corresponding order of complexity.

3.1 Mixed logit

We defined the mixed logit as any model in which probabilities can be expressed as,

$$P_{ni} = \int \left(\frac{e^{ni}}{\sum_{j=1}^{J} e^{nj}} \right) f(\beta) d\beta \tag{9}$$

However, the mixed logit can also be interpreted as a model with a correlation error term between the utilities of different alternatives. The utility can be expressed as $U_{njt} = \beta_n' x_{njt} + \epsilon_{njt}$. And the coefficients β_n can be decomposed in its mean α and its deviations μ_n, therefore the utility can be written as,

$$U_{nj} = \alpha' x_{nj} + \mu_n' x_{nj} + \epsilon_{nj} \tag{10}$$

where x_{nj} is the observable variables of alternative j, α is a vector of fix coefficients, μ_n' is a vector of random terms with mean equal to zero and ϵ_{nj} is identically distributed Extreme Value Type I. The non-observed part of the utiliy is therefore,

$$\eta_{nj} = \mu_n' x_{nj} + \epsilon_{nj} \tag{11}$$

And the correction between alternatives is: $Cov(\eta_{ni}, \eta_{nj} = E(\mu_n' x_{ni} + \epsilon_{ni})(\mu_n' x_{nj} + \epsilon_{nj}) = x_{ni} W x_{nj}$. In this way, the utility is correlated between alternatives even if the error component is independent. As a consequence it does not exhibit IIA, as the ratio between two mixed logit probabilities $\frac{P_{ni}}{P_{nj}}$ models depends on all the data, not on only of the attributes of i and j

With the mixed logit formulation, we can model panel data by specifying variability in the coefficients between different decision-makers but constant coefficients for the same individual in different situations.

The mixed logit where the coefficients β_n are distributed with a density $f(\beta|\theta)$ can be easily simulated by extracting a random value from $f(\beta|\theta)$, evaluating it in the logit formula,

$$L_{ni}(\beta) = \frac{e^{\beta x_{ni}}}{\sum_{j=1}^{J} e^{\beta x_{nj}}} \tag{12}$$

and repeating this process several times.

Finally, the simulated probability is the average of the previous values.

$$P_{ni} = \frac{1}{R} \sum_{r=1}^{R} L_{ni}(\beta^r) \tag{13}$$

3.2 Latent class choice model

LCCM are composed by two sub-models, a class membership model and a class-specific choice model which log-likelihood is maximize simultaneously.
The class membership model returns the probability of an individual n to belong to a given class k. It is dependent of its characteristics, and the utility can be expressed as follows:

$$U_{nk} = S_n\gamma_k + \upsilon_{nk} \tag{14}$$

where S_n is the vector containing the characteristics of the individual n, γ_k is the vector of the unknown parameters that need to be estimated and υ_{nk} is the error term that typically is assumed to be independently and identically distributed Extreme Value Type I over decision-makers and classes.
Therefore, the probability that the decision-maker n belongs to class k given his/her characteristic is the logit formula:

$$P(q_{nk}|S_n, \gamma_k) = \frac{e^{S_n\gamma_k}}{\sum_{k=1}^{K} e^{S_n\gamma_k}} \tag{15}$$

where $q_{nk} = 1$ if he/she belongs to class k and 0 otherwise.
On the other hand, the class-specific choice model returns the probability that the decision maker choose an specific alternative, given that she/he belong to an certain class.
The utility of individual n when choosing alternative i during the period t, given that he/she bellows to class k is formulated as:

$$U_{nit|k} = X'_{nit}\beta_K + \epsilon_{nit|k} \tag{16}$$

where X_{nit} is the vector of observed attributes of alternative i during time period t normally including an alternative-specific constant. β_k is the vector of unknown parameters that are being estimated and $\epsilon_{nit|k}$ is independent and normally distributed Extreme Value Type I over decision-makers, alternatives and classes.
Therefore, the probability that individual n chose alternative j during time period t, given that he/she belongs to class k is formulated as

$$P(y_{nit}|X_{nit}, q_{nk}, \beta_k) = \frac{e^{V_{nit|k}}}{\sum_{j=1}^{J} e^{V_{njt|k}}} \tag{17}$$

where J is the number of alternatives.
Thus, conditional on the class, the probability of observing y_n, during all time periods T_n is formulated as:

$$P(y_n|X_n, q_{nk}, \beta_k) = \prod_{t=1}^{Tn}\prod_{j=1}^{J}(P(y_{njt}|X_{njt}, q_{nk}, \beta_k))^{y_{njt}} \tag{18}$$

where y_{njt} is 1 if the decision-maker n choose alternative j during the time period t and 0 otherwise.
The unconditional probability of observing y_n, can be calculated by mixing the previous choice probability by the probability of belonging to each class k:

$$P(y_n) = \sum_{k=1}^{K} P(q_{nk}|S_n, \gamma_k) P(y_n|X_n, q_{nk}, \beta_k) \tag{19}$$

To conclude, the likelihood over all individuals N, considered independence between them, is expressed as follows:

$$P(y) = \prod_{n=1}^{N} \sum_{k=1}^{K} P(q_{nk}|S_n, \gamma_k) P(y_n|X_n, q_{nk}, \beta_k) \tag{20}$$

On the other hand, it is worth mentioning that we have employed the AIC and BIC goodness-of-fit and the corresponding variances and p-values to select the attributes and socio-demographic variables present in each of the models that we will present in section 5. We overview the employed goodness-of-fit measures for completeness in the two following subsections.

3.2.1 Likelihood ratio test

The likelihood-ratio test assesses the goodness of fit of two competing statistical models based on the ratio of their likelihoods, precisely when one is an extended model on the parameter space and the other one has some constraints imposed. It is a good measure to detect overfitting.
If the null hypothesis is true, then the two models' likelihood differences should only differ by the sampling error. The sampling error difference is modeled by a χ^2 distribution with df degrees of freedom.
Therefore, the likelihood ratio test is defined, for the null hypothesis (H_0) being true, by:

$$\chi^2_{df} \sim 2(LL(Extended) - LL(baseline)) \tag{21}$$

where the degrees of freedom df, is given by,

$$df = parameters_{Extended} - parameters_{baseline} \tag{22}$$

3.2.2 AIC and BIC

The Akaike information criterion (AIC) is a mathematical test used to evaluate how well a model fits the data. It penalizes models that use more parameters to avoid over-fitting since we aim to select the model that better explains the data with the smallest number of parameters.

$$AIC = 2k - 2ln(\hat{\mathcal{L}}) \tag{23}$$

where k is the number of parameters in the model and $\mathcal{L}$ is the estimated likelihood of the model.
On the other hand, the Bayesian information criterion is closely related to the AIC but derived from Bayesian probability, is expressed as:
.

$$BIC = klog(n) - 2ln(\hat{\mathcal{L}}) \tag{24}$$

where k is the number of parameters in the model, n is the number of observations and $\mathcal{L}$ is the estimated likelihood of the model.
For model selection purposes, there is no clear choice between AIC and BIC. Both formulations introduce a penalty term for the number of parameters in the model, but this penalty term is larger in the BIC than in the AIC. We employ both criteria for our models selection.

3.3 Latent class choice model with attitudinal statements

In this section we formulate the traditional LCCM model, where the factors scores obtained from a Confirmatory Factor Analysis (CFA) of attitudinal statements are included in the class membership model as continuous variables.
In other words, a pre-computed representation of the latent variable space of the attitudinal variables is added into the LCCM as continuos variables. However, this approach treats CFA factors as observed variables in the LCCM model and does not perform a joint estimation of classes and latent variables (factors).
Hereafter, we introduce the basic factor analysis concepts employed for this and subsequent formulations.

3.3.1 Exploratory Factor Analysis

To not excessively increase the number of model parameters, as it could make the computation excessive or overfit the model, we perform Factor Analysis to the attitudinal variables.
Factor analysis explains the variance among the observed variable and condenses a set of the observed variables into unobserved variables called factors. Observed variables are modeled as a linear combination of factors and error terms that also help understand which characteristics are strong when determining groups.
The relation between the observed variables (x_k) and the unobserved factors (F_j) is given by the equation:

$$x_k = \bar{x_k} + \sum_j \rho_{kj} F_j + \varphi_k \tag{25}$$

where $\bar{x_k}$ is the mean value of the observed variable k and φ_k is the error term model as a normal distribution. The factor loadings, ρ_{kj}, quantify the correlation between the observed variable and unobserved factors.
The Principal axis factoring method has been employed with Varimax (orthogonal) rotation.
Before computing the factor analysis, we tested our data's adequacy with several tests.

Firstly, Bartlett's Test states the null hypothesis (H_0) that the correlation matrix is an Identical matrix. So, it assumes that no correlation is present among the variables. Therefore, we expect a low p-value that contradicts the null hypothesis (statistically significant variances). Secondly, the Kaiser-Meyer-Olkin (KMO) test measures how suited the data is for Factor Analysis. It computes the proportion of variance among variables that might be common variance. The lower the proportion, the higher the KMO-value and the more suited the data is to Factor Analysis.

3.3.2 Confirmatory Factor Analysis

After computing the Exploratory Factor Analysis, which allows us to have a predefined structure on how the factors are constructed and related, we can compute the Confirmatory Factor Analysis (CFA). CFA basic assumption is that each factor is associated with a particular set of observed variables. It is a Restricted Model since we impose some restrictions on its structure, such as not having cross-loadings of a factor with variables from other factors. In other words, CFA creates independent factors and confirms the structure that the EFA predefines.
Once the CFA confirms the factor structure, we employ the Cronbach's alpha test that measures internal consistency. Cronbach's alpha test indicates how closely related a set of items are as a group. It is a weighted average of the correlations between the variables that form part of the group. Therefore, the closer it is to 1, the more consistent the items will be with each other. Ideally, it should be at least 0.7, but we can also accept factors with a lower level if they have a coherent meaning.

3.4 Gaussian-Bernoulli Mixture Latent Class Choice Model

We followed the model formulation presented in [31] for our next model.
In Gaussian-Bernoille Mixture Latent Class Choice Model, the class membership model, $P(q_{nk}|S_n, \gamma_k)$ is defined as a Gaussian-Bernoulli Mixture Model (GBM), a probabilistic machine learning technique used for clustering, where the Gaussian Mixture Model (GMM) is used for continuous variables and the Bernoulli Mixture Model (BMM) for binary variables. In this way, the characteristics of the decision- maker (S_n) are divided in S_{cn}, that refers to the continuous characteristic and has a dimension D_c corresponding with the number of elements in S_{cn}, and in S_{dn} that accounts for the discrete/binary characteristics of individual n and has a dimension D_d corresponding with the number of elements in S_{dn}.
The Gaussian Mixture model is defined as a combination of normal distributions $\mathcal{N}(S_{cn}|\mu_{ck}, \Sigma_{ck})$ where each component has its own mean μ_{ck} (with dimension equal to the number os elements in S_{cn}), covariance Σ_{ck} and mixing coefficient π_k. On the other hand, BMM is a combination of k mixture components, each being the product of D_d independent Bernoulli probability distributions with its own mean vector μ_{dk}.
However, the formulation of the class membership as a GBM cannot be directly applied into the model. The probability of a decision-maker n belonging to class k, given its characteristics, $P(q_{nk}|S_n)$ is the posterior probability calculated employing Bayes' therorem, and

therefore, cannot appear in the maximization of the likelihood function.
Instead, we estimate the probability of observing decision-maker n with characteristics $S_n = \{S_{cn}, Sdn\}$ given that he/she belongs to latent class k. The model architecture is represented in Figure 1, where unobserved variables are represented by ellipses and observable variables by rectangles.
By making the assumption that the continuous and binary data of the Gaussian and Bernoulli distributions are independent, the joint probability can be expressed as as the product of the probability of the class, the densities of S_{cn} and S_{dn} given the class, and the choice probability conditional on the class:

$$P(S_{cn}, S_{dn}, y_n, q_{nk}) = P(q_{nk}|\pi_k)P(S_{sn}|q_{nk} = 1, \mu_{ck}, \Sigma_{ck})P(S_{dn}|q_{nk} = 1, \mu_{dk})P(y_n|X_n, q_{nk}, \beta_k) \tag{26}$$

where

$$P(q_{nk}|\pi_k) = \pi_k \tag{27}$$

$$\sum_{k=1}^{K} \pi_k = 1 \tag{28}$$

π_k is the probability of beloging to class k regardless the characteristics of the decision-maker, also called prior probability. Its sum over all the classes is equal to 1.
The densities of S_{cn} and S_{dn}, given the class, are expressed as:

$$P(S_{cn}|q_{nk} = 1, \mu_{ck}, \Sigma_{ck}) = \mathcal{N}(S_{cn}|\mu_{ck}, \Sigma_{ck}) = \frac{1}{\sqrt{(2\pi)^{D_c}|\Sigma_{ck}|}} exp\left(-\frac{1}{2}(S_{cn} - \mu_{ck})'\Sigma_{ck}^{-1}(S_{cn} - \mu_{ck})\right) \tag{29}$$

$$P(S_{dn}|q_{nk} = 1, \mu_{dk}) = \prod_{i=1}^{D_d} \mu_{dk_i}^{S_{dni}}(1 - \mu_{dk_i})^{(1-S_{dni})} \tag{30}$$

where $|\Sigma_{ck}|$ is the determinant of the covariance matrix, S_{dni} is a binary characteristics of decision-maker n and μ_{dki} its corresponding mean.
Therefore, the joint probability of S_{cn}, S_{dn} and y_n can be calculated by taking the marginal of (26) over the K classes.

$$P(S_{cn}, S_{dn}, y_n) = \sum_{k=1}^{K} P(S_{cn}, S_{dn}, y_n, q_{nk}) \tag{31}$$

Finally, the likelihood function of the complete model for all the decision makers N is expressed as:

$$\begin{gathered} P(S_c, S_d, y) = \prod_{n=1}^{N} P(S_{cn}, S_{dn}, y_n) = \\ \prod_{n=1}^{N}\sum_{k=1}^{K} \pi_K \mathcal{N}(S_{cn}|\mu_{ck}, \Sigma_{ck}) \prod_{i=1}^{D_d} \mu_{dk_i}^{S_{dni}}(1 - \mu_{dk_i})^{(1-S_{dni})} \prod_{t=1}^{T_n}\prod_{j=1}^{J}\left(\frac{e^{X'_{njt}\beta_k}}{\sum_{j'=1}^{J} e^{X'_{nj't}\beta_k}}\right)^{y_{njt}} \end{gathered} \tag{32}$$

Typically, discrete choice models are estimated using the maximum likelihood estimation to maximize the observed data's likelihood given the model parameters. However, maximizing the LCMM and the GBM-LCCM together is complex due to the summation over the k classes that appears in the final likelihood and will not lead to a closed-form solution.
To solve this problem, we apply the Expectation-Maximization (EM), a commonly used algorithm to maximize the likelihood estimation in models with latent classes.

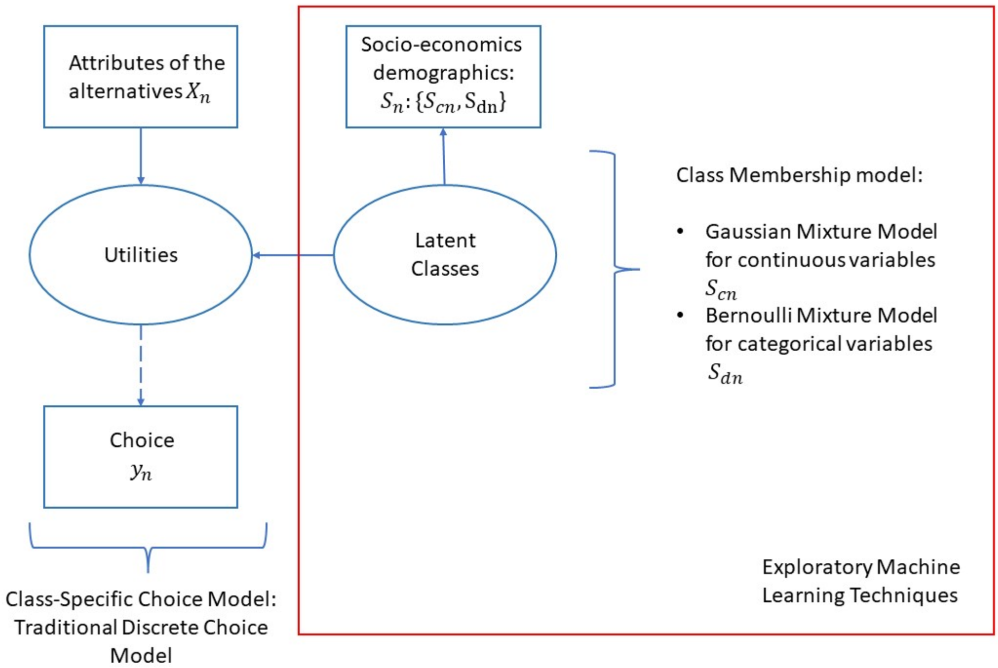

Figure 1: Model struture of Gaussian-Bernoulli Mixture Latent Class Choice Model (GBM-LCCM)

3.4.1 EM algorithm

The EM algorithm is a method that combines an expectation step with a maximization one until we reach convergence. We start by initializing the unknown parameters. After, we estimate the expected values of the latent variables (E-step) using Bayes' theorem. Then, we update the unknown parameters' values, using the maximization of the log-likelihood (M-step).
Finally, we evaluate the log-likelihood with the current values of the unknown parameters and check for convergence. We return to the E-step, until convergence is reached.

We can rewrite the joint likelihood assuming that the clusters are observed as follows:

$$\begin{aligned} P(S_c, S_d, y, q) = &\prod_{n=1}^{N}\prod_{k=1}^{K}\left[\pi_K \mathcal{N}(S_{cn}|\mu_{ck}, \Sigma_{ck})\prod_{i=1}^{D_d}\mu_{dk_i}^{S_{dni}}(1-\mu_{dk_i})^{(1-S_{dni})}\right]^{q_{nk}} \\ &X\prod_{n=1}^{N}\prod_{k=1}^{K}\prod_{t=1}^{T_n}\prod_{j=1}^{J}\left[\frac{e^{X'_{njt}\beta_k}}{\sum_{j'=1}^{J}e^{X'_{nj't}\beta_k}}\right]^{y_{njt}q_{nk}} \end{aligned} \tag{33}$$

The function is divided into two parts when we take the logarithm of the likelihood:

$$\begin{aligned} LL = &\sum_{n=1}^{N}\sum_{k=1}^{K} q_{nk} log\left[\pi_K \mathcal{N}(S_{cn}|\mu_{ck}, \Sigma_{ck})\prod_{i=1}^{D_d}\mu_{dk_i}^{S_{dni}}(1-\mu_{dk_i})^{(1-S_{dni})}\right] \\ &+\sum_{n=1}^{N}\sum_{k=1}^{K}\sum_{t=1}^{T_n}\sum_{j=1}^{J} y_{njt}q_{nk} log\left[\frac{e^{X'_{njt}\beta_k}}{\sum_{j'=1}^{J}e^{X'_{nj't}\beta_k}}\right] \end{aligned} \tag{34}$$

this equation can be solved by equating to zero its derivatives with respect to each of the unknown parameters only if q_{nk} is known.
In order to find the initial values of q_{nk}, we estimate the expectation of q_{nk} (E-step) using Bayes's theorem.

$$\begin{aligned} &E[q_{nk}] = \gamma_{qnk} = \\ &= \frac{\pi_K \mathcal{N}(S_{cn}|\mu_{ck}, \Sigma_{ck})\prod_{i=1}^{D_d}\mu_{dk_i}^{S_{dni}}(1-\mu_{dk_i})^{(1-S_{dni})}\prod_{t=1}^{T_n}\prod_{j=1}^{J}\left[\frac{e^{X'_{njt}\beta_k}}{\sum_{j'=1}^{J}e^{X'_{nj't}\beta_k}}\right]^{y_{njt}}}{\sum_{k'=1}^{K}\left[\pi_{K'} \mathcal{N}(S_{cn}|\mu_{ck'}, \Sigma_{ck'})\prod_{i=1}^{D_d}\mu_{dk'_i}^{S_{dni}}(1-\mu_{dk'_i})^{(1-S_{dni})}\prod_{t=1}^{T_n}\prod_{j=1}^{J}\left[\frac{e^{X'_{njt}\beta_{k'}}}{\sum_{j'=1}^{J}e^{X'_{nj't}\beta_{k'}}}\right]^{y_{njt}}\right]} \end{aligned} \tag{35}$$

where γ_{qnk} is considered as the posterior probability of the classes. Combining equations (34) and (35), gives:

$$\begin{aligned} E[LL] = &\sum_{n=1}^{N}\sum_{k=1}^{K} \gamma_{nk} log\left[\pi_K \mathcal{N}(S_{cn}|\mu_{ck}, \Sigma_{ck})\prod_{i=1}^{D_d}\mu_{dk_i}^{S_{dni}}(1-\mu_{dk_i})^{(1-S_{dni})}\right] \\ &+\sum_{n=1}^{N}\sum_{k=1}^{K}\sum_{t=1}^{T_n}\sum_{j=1}^{J} y_{njt}\gamma_{nk}\left[\frac{e^{X'_{njt}\beta_k}}{\sum_{j'=1}^{J}e^{X'_{nj't}\beta_k}}\right] \end{aligned} \tag{36}$$

where the derivatives of the expected log-likelihood with respect to the unknown parameters can be settled to zero. The following expressions are obtained for each parameter:

$$\mu_{ck} = \frac{1}{N_k}\sum_{n=1}^{N}\gamma_{q_{nk}} S_{cn} \tag{37}$$

$$\Sigma_{ck} = \frac{1}{N_k}\sum_{n=1}^{N}\gamma_{q_{nk}}(S_{cn}-\mu_{ck})(S_{cn}-\mu_{ck})' \tag{38}$$

$$\mu_{dk} = \frac{1}{N_k}\sum_{n=1}^{N}\gamma_{q_{nk}}S_{dn} \tag{39}$$

$$\pi_k = \frac{N_k}{N} \tag{40}$$

$$\beta_k = argmax_{\beta_k}\sum_{n=1}^{N}\sum_{t=1}^{Tn}\sum_{j=1}^{J} y_{njt}\gamma_{q_{nk}} log\left[\frac{e^{X'_{njt}\beta_k}}{\sum_{j'=1}^{J} e^{X'_{nj't}\beta_k}}\right] \tag{41}$$

where N_k is defined as:

$$N_k = \sum_{n=1}^{N}\gamma_{q_{nk}} \tag{42}$$

It is worth mentioning that no closed-form solution can be obtained for the case of beta parameters (equation 41). Thus, we employ the gradient-based numerical optimization method BFGS (Nocedal et al., 1999).
After convergence is reached, we can calculate the probability of observing a vector of choices y of all decision-makers as follows:

$$P(y) = \prod_{n=1}^{N}\sum_{k=1}^{k} P(q_{nk}|S_{cn}, S_{dn}, \mu_{ck}, \Sigma_{ck}, \mu_{dk}, \pi_k)\prod_{t=1}^{T_n}\prod_{j=1}^{J}(P(y_{njt}|X_{njt}, q_{nk}, \beta_k))^{y_{njt}} \tag{43}$$

where $P(q_{nk}|S_{cn}, S_{dn}, \mu_{ck}, \Sigma_{ck}, \mu_{dk}, \pi_k)$ is the posterior probability of vector $S_n = \{S_{cn}, S_{dn}\}$ being generated by cluster k using Bayes' theorem:

$$P(q_{nk}|S_{cn}, S_{dn}, \mu_{ck}, \Sigma_{ck}, \mu_{dk}, \pi_k) = \frac{P(q_{nk}|\pi_k)P(S_{cn}|q_{nk}, u_{ck}, \Sigma_{ck})P(S_{dn}|q_{nk}, \mu_{dk})}{\sum_{k'=1}^{K} P(q_{nk'}|\pi_{k'})P(S_{cn}|q_{nk'}, u_{ck'}, \Sigma_{ck'})P(S_{dn}|q_{nk'}, \mu_{dk'})} \tag{44}$$

3.5 Gaussian-Bernoulli Mixture Latent Class Choice Model with attitudinal statements

Based on the formulation of the GBM-LCCM model presented in section 3.4, we extended the model by including the information of the attitudinal questions.
In this case, the definition of the latent classes includes the factor scores obtained from the Confirmatory Factor Analysis (CFA) of the attitudinal variables, as we made in section 3.3.2. This approach also treats CFA factors as observed variables in the LCCM model and does not perform a joint estimation of classes and latent variables (factors).
The model structure is presented in Figure 2, where unobserved variables are represented by ellipses and observable variables by rectangles.

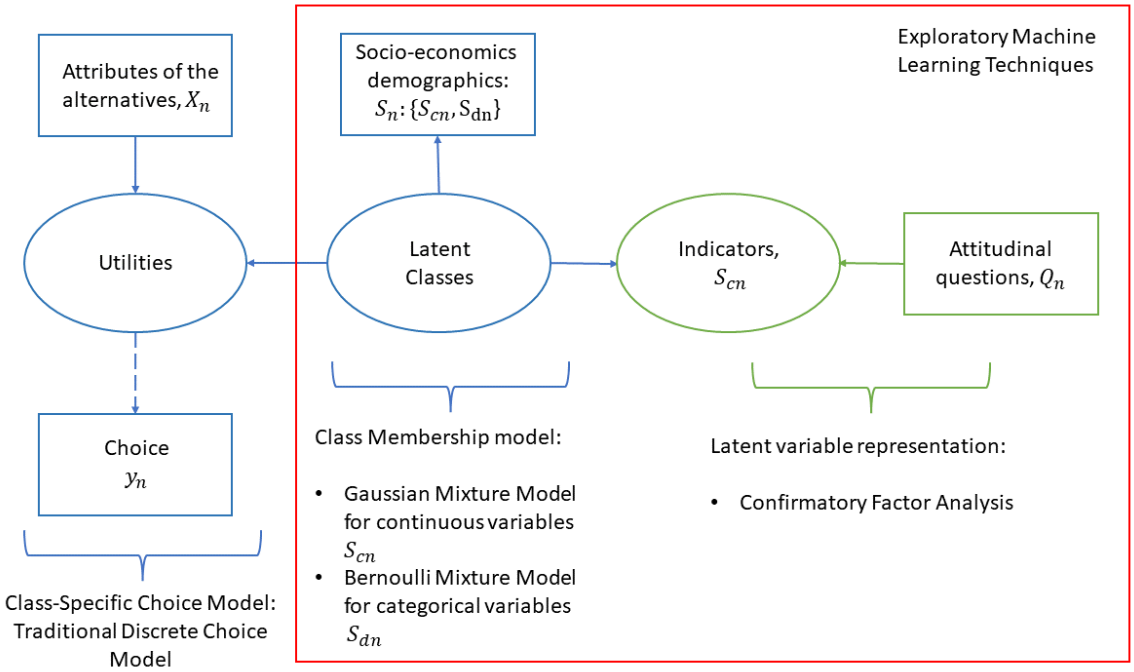

Figure 2: Model structure of Gaussian-Bernoulli Mixture Latent Class Choice Model (GBM-LCCM) with attitudinal statements

In figure 2, we observed how the psychological indicators, in this case, the factor scores results from the CFA, are included in the class membership model. The colors are different to indicate that the indicators are included as pre-computed information into the model, and no joint estimation is performed.

3.6 Embeddings model formulation

3.6.1 Interpretable Embeddings MNL (E-MNL)

The utility $U_{i,n}$ that individual n associates with alternative i is expressed as:

$$U_{i,n}(X_{i,n}, Q_{i,n}) = V_{i,n}(X_{i,n}, Q_{i,n}) + \epsilon \tag{45}$$

where ϵ is independent and normally distributed Extreme Value Type I, $X = \{X_1, X_2, ..., X_k\}$ is the set of continuous explanatory variables and $Q = \{Q_1, Q_2, ..., Q_M\}$ is the set of categorical explanatory variables. Both can describe observed attributes of the alternatives, individual's socio-demographic or attitudinal variables.
Employing a ANN-based DCM model, the systematic part of the utility $V_{i,n}$ is formulated as:

$$V_{i,n}(X_n, Q_n) = f_{i,1}(X_n; B) + f_{i,2}(g_i(Q_n; W_i); B')) \tag{46}$$

where g represents the embedding function, that maps each dummy encoding input Q_n to a latent J-dimensional representation Q'_n based on a set of trainable weights, W. The shape

of W is ZxJ, where Z is the number of unique categories, $Z = \sum_{m=1}^{M} |Q_m|$ and J is the number of alternatives in the choice set C. Matrix-multiplication is performed between Q and W resulting in Q', which shape is MxJ.
On the other hand, the function f is defined as a linear function, such that the trainable preference parameters of the model B and B' are a linear combination of the explanatory variables: $f_{i,1}(X_n; B) = BX_{i,n}$ and $f_{i,2}(g_i(Q_n; W_i); B') = B'Q'_{i,n}$.
The model architecture is shown in Figure 3. The ANN consist of two input layers receiving two separated inputs: X and Q. X is directly connected to the first convolution filter with the set of trainable weights B. On the other hand, the one-hot encoded inputs Q, are projected into the embedding layer and then, the new alternative-specific representation Q' is connected to the second convolution filter with the set of trainable weights B'
The final likelihood of individual n selecting choice alternative i is formulated using the softmax function, as:

$$P_n(i) = \frac{e^{f_{i,1}(X_n;B)+f_{i,2}(g_i(Q_n;W_i);B'))}}{\sum_{j\in C_n} e^{f_{j,1}(X_n;B)+f_{j,2}(g_j(Q_n;W_j);B'))}} \tag{47}$$

which corresponds to the logit probability.[27]

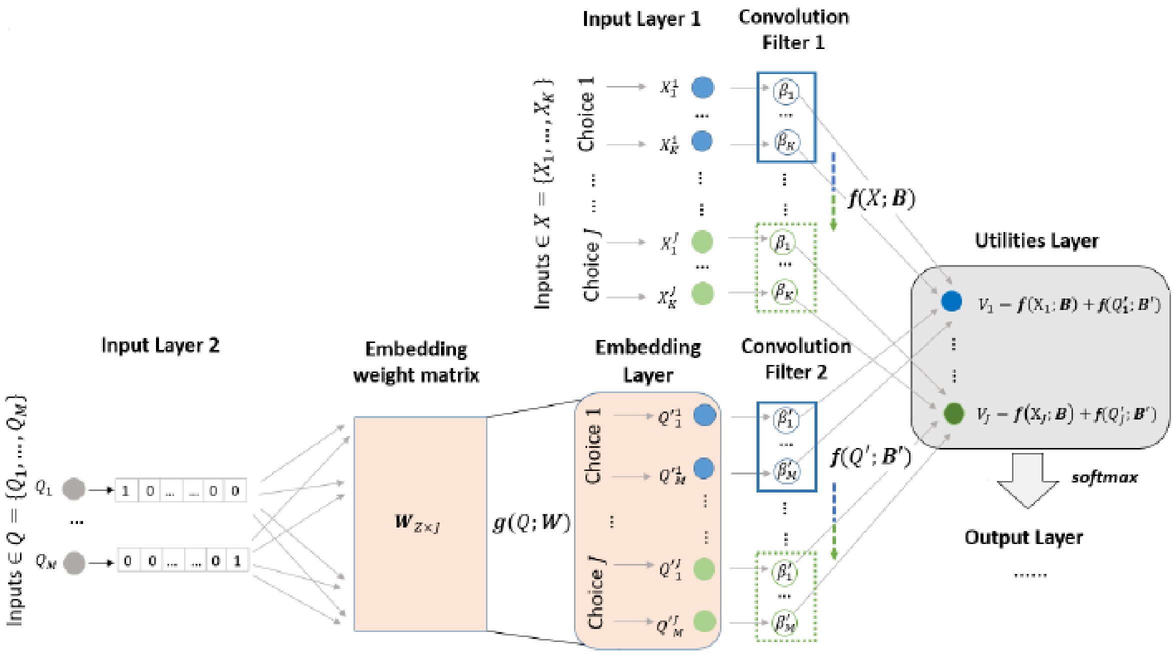

Figure 3: E-MNL model architecture. Source: [27]

3.6.2 Extended Model: Interpretable Embeddings MNL with Representation Learning term (EL-MNL)

In this last model formulation, we present the Embedding Learning Multinomial Logit (EL-MNL) that combines the previous E-MNL model with the Learning Multinomial Logit

(L-MNL) suggested by Sifringer et. al in [29]
EL-MNL extends the previous model formulation by increasing the dimensionality of the embeddings matrix (W) from J to $J + S$. The extra dimensions S of Q' are the inputs to a densely connected hidden layer L with K hidden nodes $q_1, ..., q_K$. The output of L is a J-dimensional vector r, where each single term (r_i) corresponds with an alternative.
In other words, r_i, also called in [29] 'representation learning term' is added to each of the utility functions to reduce the problems of endogeneity and correct for underfit due to undetected misspecification of the disturbance term, increasing the overall predictive performance compared to the E-MNL model.
Therefore, the systematic part of the utilities $V_{i,n}$ are expressed as:

$$\begin{aligned} V_{i,n}(X_n, Q_n) = f_{i,1}(X_n; B) + f_{i,2}(g_i(Q_n; W_i); B'))+ \\ r_i(\{g_{J+1}(Q_n; W_{J+1}), ..., g_{J+S}(Q_n; W_{J+S})\}; M_i, \alpha_i) \end{aligned} \tag{48}$$

where the functions f, g and the trainable weights B', B, W are the same as define in 3.6.1, but the size of W in this case is MxD, $D = J + S$. The function r represents the operations performed in the densely connected layer L that receives $R_n = \{Q'_{J,n}, ..., Q'_{J+S,n}\}$ as input.
Therefore, (48) can be expressed as:

$$V_{i,n}(X_n, Q_n) = f_{i,1}(X_n; B) + f_{i,2}(Q'_{i,n}; B')) + r_i(R_n; M_i, \alpha_i) \tag{49}$$

where the function r is define as:

$$r_i(R_n; M_i, \alpha_i) = \sum_{m \in M_i} k(max(0, R_n)) + \alpha_i \tag{50}$$

being M_i an alternative-specific trainable parameters and α_i a bias term.
The model architecture is presented in figure 4.
In [29] it is proof that if the set of features that enters the function f do not overlap with the ones entering the function r, the parameters of f can still be interpretable. As shown in 48, the constraint is met for this formulation and, interpretability is still maintained for some input variables. However, it is not fully interpretable as the E-MNL model.[27]

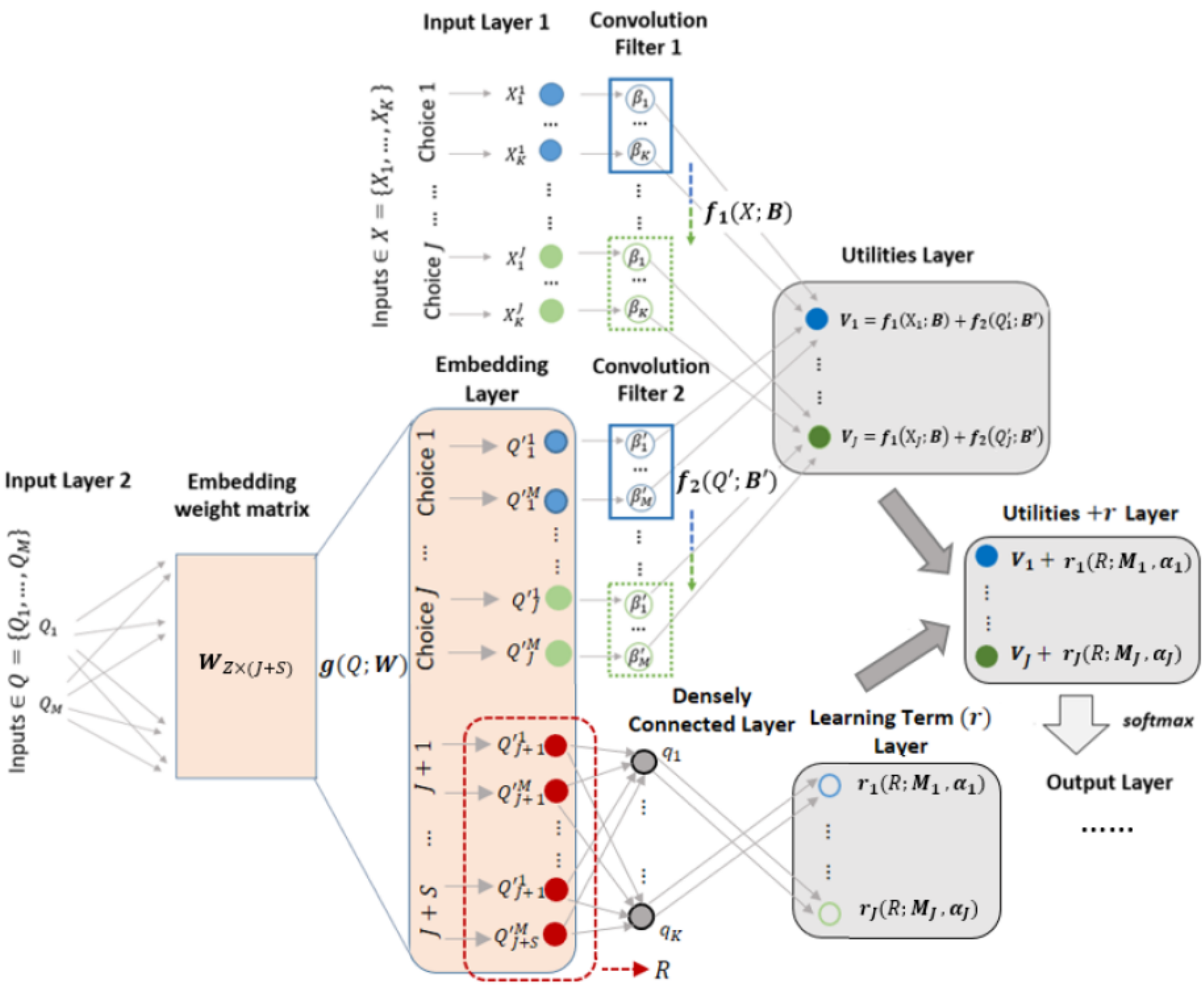

Figure 4: EL-MNL model architecture. Source: [27]

3.6.3 Interpretability

By following the forehead formulations of the embeddings, we can obtain unique insights into how the categorical variables influence each of the choice alternatives. In addition to providing an encoding method to convert categorical variables to continuous.
Because the dimension of the embedding matrix is fixed and is not subject to tuning, we obtain interpretable embeddings where the value of the embedding along to the j-th dimension represent the relevance of the encoded category to the j-th choice alternative. In the case of the EL-MNL model of section 3.6.2, the first J dimensions remain interpretable correspoding with the J choice alternatives, as in the E-MNL formulation.
For interpreting the embeddings, we need to pay attention to the sign and value each category receives along each dimension of the embedding space and the relative distances between every pair of points. By restricting the coefficients associated to Q' to be positive, we en-

hance the interpretability of the emdeddings values, such that a high and positive value along an alternative-specific axis shows a high positive effect of the selected category on the corresponding choice alternative. Moreover, closer proximity between categories in the embedding space demonstrates that their overall impact on the alternatives is similar. [27]

4 Data

The data used in this study comes from the Share-More project, which aims to maximize the value of car-sharing services to encourage sustainable urban mobility by developing a framework of personalized incentives.
The data was collected through a tailor-made online survey (available both web and mobile versions), from July 16^{th} to August 6^{th} 2020 and simultaneously in three cities: Copenhagen, Munich, and Tel Aviv. The survey was available in English and Danish for Copenhagen (CPH) and its duration was estimated at around 15 minutes. The sample was opportunistic, with 200 individuals per city as a target sample size. CPH and Munich's respondents were recruited through panels, while in Tel Aviv-Yafo, they were contacted through different mailing lists. The eligibility criteria were more than 18 years old and having a valid driver's license. In Tel Aviv, the minimum age for using car-sharing services is 21 years; therefore, the eligibility criteria in this city was having 21 years or more.
The survey was divided into six parts. It started with a brief introduction to the project and its objective. The second part consisted of socio-demographic questions, including experience with car-sharing. In the next section, the survey addressed questions regarding the responders' attitude toward private and car-sharing. The fourth part was questions related to preferences about car-sharing incentives. The fifth part was a Stated Preference(SP) experiment in which different options of car-sharing plans, taking into account existent services in each city. The last part of the survey was to examine the effect of Covid-19 on respondents' mobility behavoir.
For our analysis, only data from Copenhagen and parts 2, 4, and 5 of the questionnaire will be employed. For further details on the survey and sample, the reader is referred to [1]

4.1 Copenhagen's mobily context

The context in which the data is collected is vital for its analysis. Therefore, Copenhagen mobility context is presented hereafter.
Copenhagen metropolitan area has a population of 1.846.023 inhabitants with an area of 2.562,80 Km^2, which leads to a density of 720,31 hab/km^2. However, when only taking into account the city of Copenhagen, the density achieves a value of 7.455,92 hab/km^2.
On the other hand, Copenhague's mobility landscape is quite particular. Cycling is a common way of getting around. The city contains a complete network of dedicated bicycle infrastructure that has 382 km and it is possible to board the metro, train and harbor bus with bikes. As a result, 49% of the population commute by bike to work or education and 63 % of school kind bike or walk to school. Share bikes and e-bikes are also available in Copenhagen.

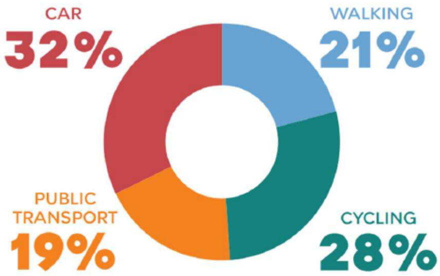

Figure 5: Modal share of Copenhagen considering trip to, from and in Copenhagen in 2018
Source: City of Copenhagen, 2019

In 2018, Copenhagen's modal share revealed that 49% of all trips are made by soft modes, 19% by public transport, and 32% by car.
The public transport in Copenhagen consists of buses, harbor buses, a driverless metro, and an urban-suburban rail (S-train) that serves the Greater Copenhagen. They are all under a standard system for fare zones and tickets at a national level.
Regarding taxation for car ownership, the registration tax accounts for 85% of the taxable value of the car up to DKK 185.100. For those above DKK 185.100, taxes are 150% of the taxable value of the car. In 2021, electric cars only have a tax of 65% of the calculated vehicle registration tax. In 2022, the price to pay will be 90% of the calculated vehicle registration tax.
The availability of car-sharing services is increasing in Copenhagen during the last decade. One hundred ninety-two parking spaces are reserved for station-based car-sharing cars, 7% of those destined for electric cars. The most common car-sharing companies in CPH are ShareNow and Green Mobility, both free-floating services, and GoMore, a peer-to-peer car-sharing company. These companies have been operating in the city for around five years.

4.2 Sample characteristics

A total of 543 Copenhagen inhabitants answered the survey. However, inconsistent respondents were deleted. They were those who stated to be aware of car-sharing, and, later in the survey, they answered the opposite. Also, respondents who finished the survey in less than 40% of the median time were removed, suggesting a lack of attention. After the cleaning, the sample size of Copenhagen consisted of 542 respondents.
The sample characteristics are describe in table 1. More than 90% of respondents stated to be aware of car-sharing services, which indicates that car-sharing services are well-known. The sample is quite balanced regarding gender and proportionally representative of the population in terms of age. Most of the respondents live in the city center and are employed. For the level of education, more than 60% of the population have at least a bachelor's degree. Moreover, most respondents live in households of 1 or 2 members and up to one car.

Additionally, more than 50 percent of the sample have two or more bikes at home.
Given the income level, most respondents earn around average (350.000 kr./year) or above, which is likely related to the high level of education of the sample.

		Total	%
Car -sharing membership status	Car-sharing member	96	17.68
	Past-car sharing member	64	11.79
	Non-car-sharing member	383	70.53
Car-sharing awareness	Yes	490	90.24
	No	53	9.76
Gender	Man	267	49.17
	Woman	275	50.64
	Prefer not to say	1	0.18
Age	18-30	146	26.89
	31-40	88	16.21
	41-50	97	17.86
	51-60	88	16.21
	More than 60	124	22.84
Place of residence	City center	235	43.28
	Suburbs	190	34.99
	Another city in the metropolitan region	71	13.08
	Outside the metropolitan region	47	8.66
Employment status	Student	74	13.63
	Employed	354	65.19
	Unemployed	12	2.21
	On leave	7	1.29
	Retired	100	18.42
	Other	8	1.47
Level of education	Less than high school	39	7.18
	High school diploma or equivalent	150	27.62
	Bachelor's degree	169	31.12
	Master's degree	134	24.68
	Doctoral degree	8	1.47
	Other	17	3.13
	Did not answer	26	4.79
Size of the household	1	152	27.99
	2	223	41.07
	3	80	14.73
	4	68	12.52
	>4	20	3.68
Number of cars in the household	0	139	25.60
	1	304	55.99
	2	91	16.76
	>2	9	1.66
Number of bicycles in the household	0	40	7.37
	1	128	23.57
	2	156	28.73
	>2	219	40.33
Income	Low (Up to 250.000 kr.)	82	15.1
	Medium (251-500.000 kr.)	140	25.8
	High (Over 500.000 kr.)	221	40.7
	Did not answer	100	18.4

Table 1: Copenhagen Sample characteristics

4.3 Attitudinal statements

In order to better understand believes and attitudes toward car-sharing and car ownership, respondents answered statements graded on a 5-point Likert scale from 1='strongly disagree' to 5='strongly agree'.
In figures 6 and 7, we have investigated the proportion of respondents who have selected 'agree' or 'strongly agree' at the corresponding statements.
Car ownership is usually considered a significant expense rather than a status symbol between the respondents from Copenhagen. Almost half of the sample also states that it is easier for them to conduct their daily trips without a private car, and around 40% empathize with the difficulty of finding parking slots. However, around 60% of the sample agree that driving a car is the more convenient and easy way to move around. Surprisingly, only 28% of the respondents express their concern about the environmental footprints of cars.

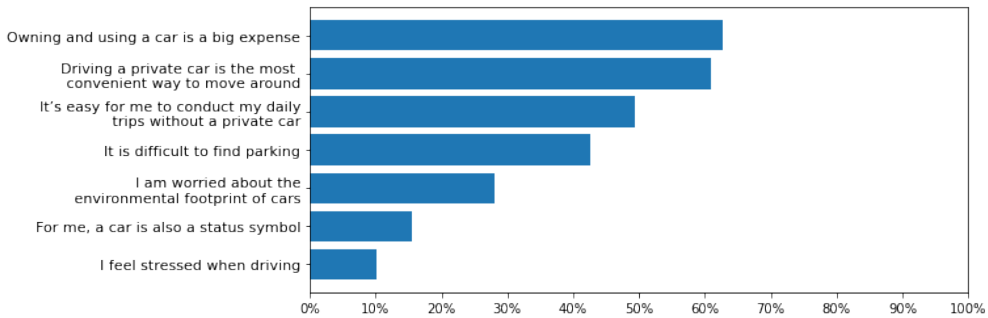

Figure 6: Sample results on attitudinal statements toward car ownenship

Regarding respondents' attitudes towards car-sharing, 67% of them agree that they do not have to deal with vehicle maintenance and repair by employing car-sharing. Moreover, up to 50% of them agree that car-sharing also saves expenses associated with private vehicle ownership. It is generally seen as a more affordable alternative than car ownership. Finally, car-sharing is also considered a more environmentally friendly alternative for most of the sample.

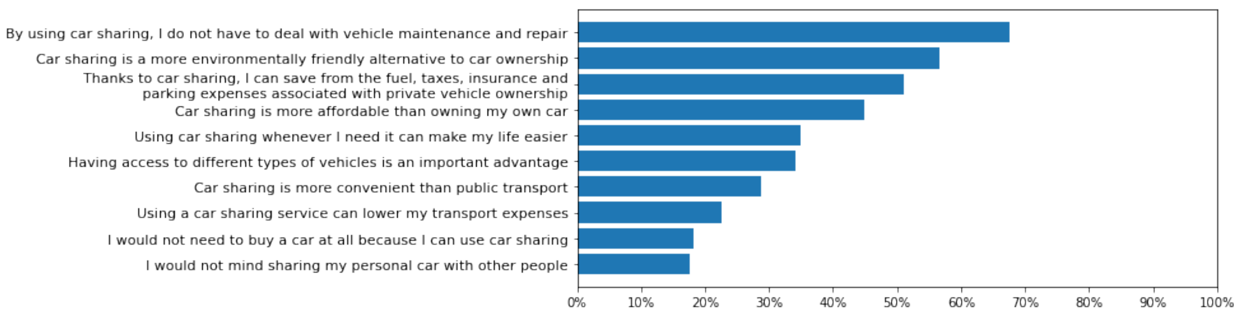

Figure 7: Sample results on attitudinal statements toward car-sharing

4.3.1 Exploratory Factor Analysis

We have computed Exploratory Factor Analyses to reduce the attitudinal variables' dimensionality.
Before computing factor analysis, we tested our data's adequacy with several tests. Firstly, Bartlett's Test states the null hypothesis that the variances are equal for all samples. Therefore, we expect a low p-value that contradicts the null hypothesis (statistically significant variances) - our analysis yields a p-value of 0. The null hypothesis is rejected, and the data seems appropriate for Factor Analysis.
Secondly, the Kaiser-Meyer-Olkin (KMO) test also measures how suited the data is for Factor Analysis. It computes the proportion of variance among variables that might be common variance. The lower the proportion, the more suited the data is to Factor Analysis. The KMO test returns a value between 0 and 1. We obtained 0.845, which can be considered an acceptable value for the computation of Factor Analysis.
The Principal axis factoring method has been employed with Varimax (orthogonal) rotation. Several exploratory factor analyses have been computed, with different numbers of factors, and even considering attitudes from CS and car ownership separately. Table 2 presents the loadings of the first four factors. For interpretation, we only considered the loadings bigger than 0.4, as advised in [33]. For the variables with the loadings in red, even if some loadings are higher than 0.4, the absolute difference between its values for different factors is less than 0.2; therefore, we do not include them.

	F1	F2	F3	F4
I feel stressed when driving.	-0,01556	0,18107104	0,432393	0,121489
For me, a car is also a status symbol.	0,005298	-0,02195947	0,037037	-0,27094
It is difficult to find parking.	0,038755	0,13902097	0,637288	-0,07415
It's easy for me to conduct my daily trips without a private car.	0,150664	0,11499914	0,379485	0,495048
Owning and using a car is a big expense.	0,138268	0,05572186	0,453027	0,111462
Driving a private car is the most convenient way to move around.	0,022581	-0,31606836	-0,32671	-0,6072
I am worried about the environmental footprint of cars.	0,232956	0,26503259	0,236902	0,208137
Car sharing is more affordable than owning my own car	0,533812	0,2689488	0,221027	0,186547
Car sharing is more convenient than public transport	0,427489	0,33290292	-0,14865	-0,37706
Car sharing is a more environmentally friendly alternative to car ownership	0,51841	0,18646079	0,225027	0,142418
By using car sharing, I do not have to deal with vehicle maintenance and repair	0,612128	0,03100724	0,140886	0,069241
Thanks to car sharing, I can save from the fuel, taxes, insurance and parking expenses associated with private vehicle ownership	0,814683	0,04983342	0,093689	0,069872
Having access to different types of vehicles is an important advantage	0,440513	0,17104531	-0,0382	-0,11046
Using a car sharing service can lower my transport expenses	0,470412	0,21612767	-0,05818	-0,10464
Using car sharing whenever I need it can make my life easier	0,379355	0,5246601	0,244928	0,025446
I would not need to buy a car at all because I can use car sharing	0,263015	0,69842185	0,291961	0,135605
I would not mind sharing my personal car with other people	0,162394	0,41872184	0,170862	0,183712

Table 2: Results of the Explanatory Factor Analysis - Loadings

After the different EFA computations and with an overview of the importance of each variable, we have tested our final structure with the confirmatory factor analysis.

4.3.2 Confirmatory Factor Analysis

Confirmatory Factor Analysis (CFA) basic assumption is that each factor is only associated with a particular set of observed variables.
We have identified the following four factors:

- Factor 1: 'Car-related issues': It refers to the inconveniences that can appear with car ownership.
- Factor 2: 'Driving a private car is the most convenient way to move around.'
- Factor 3: 'Positively inclined toward the car-sharing concept': People with a high score in this factor have a positive vision toward car-sharing. They believe that car-sharing is cheaper, easy to maintain, and more environmentally friendly than car ownership.
- Factor 4: 'Car-sharing is a good alternative to car ownership': this factor accounts for the statements which state that using car-sharing can replace the ownership of a private car specially if you do not mind sharing your car.

Even factor 4 only has one loading, we consider it sufficiently uncorrelated with the other factors and decide to include it not to lose valuable information.
Table 3 present the final loadings associated with each factor. We also computed the Cronbach's alpha test, that indicates how closely related a set of items are as a group. Ideally, it should be at least 0.7, but we also accepted the factors with a lower level since its structure is coherent with the meaning of the factors.

	F1 (Cronbach's alpha= 0.516)	F2	F3 (Cronbach's alpha= 0.578)	F4 (Cronbach's alpha= 0.578)
I feel stressed when driving.	**0.779**	0	0	0
For me, a car is also a status symbol.	0	0	0	0
It is difficult to find parking.	**1.039**	0	0	0
It's easy for me to conduct my daily trips without a private car.	0	0	0	0
Owning and using a car is a big expense.	**0.717**	0	0	0
Driving a private car is the most convenient way to move around.	0	**1.092**		0
I am worried about the environmental footprint of cars.	0	0	0	0
Car sharing is more affordable than owning my own car	0	0	0	0
Car sharing is more convenient than public transport	0	0	0	0
Car sharing is a more environmentally friendly alternative to car ownership	0	0	**0.858**	0
By using car sharing, I do not have to deal with vehicle maintenance and repair	0	0	**0.878**	0
Thanks to car sharing, I can save from the fuel, taxes, insurance and parking expenses associated with private vehicle ownership	0	0	**1.037**	0
Having access to different types of vehicles is an important advantage	0	0	**0.815**	0
Using a car sharing service can lower my transport expenses	0	0	**0.837**	0
Using car sharing whenever I need it can make my life easier	0	0	0	0
I would not need to buy a car at all because I can use car sharing	0	0	0	**0.946**
I would not mind sharing my personal car with other people	0	0	0	**0.909**

Table 3: Results of the Confirmatory Factor Analysis - Factor Loadings

4.4 State preference experiment

The fifth part of the survey was a State Preference (SP) experiment, where the results can help us better understand the effect of different car-sharing plans on the choice of the

respondents.
Each experiment task presented four car-sharing alternative plans with eight attributes and the option to select 'None of the alternatives.' The fourth task combined the previous alternatives chosen in the three experiments before. Therefore, they could choose their preferred alternative among all the ones presented to them.
Figure 21 shows an example of the format of a choice task presented to the decision-makers. To minimize response bias, the order of appearance of the attributes was random for each individual but was the same for the three tasks presented to the same individual.

? Below are presented several car sharing products with different service characteristics. Which one of these would you choose to **subscribe**? Consider that you are inside the coverage area of all services and that there are no fees other than the ones presented.

	One-way car sharing (free-floating)	**Roundtrip car sharing**	**One-way car sharing (station-based)**	**Peer-to-peer car sharing**	**None of the alternatives**
One-time subscription cost	Free	200 kr.	200 kr.	Free	
Probability to get a shared vehicle	10 out of 10 requests	9 out of 10 trip requests	10 out of 10 requests	9 out of 10 trip requests	
Extra features	Family/friends account with discounted rates	Collect credits to redeem for goods (e.g.: clothing and grocery discounts)	Guaranteed child car seat availability	Plan including other modes for a seamless door to door trip	
Walking time from parking location to destination	6 to 10 min	up to 5 min	6 to 10 min	up to 5 min	
Car sharing vehicle engine type	Electric	Electric	Electric	Electric	
Usage cost	850kr./day	500kr./day	300kr./day	800kr./day	
Car sharing vehicle types	Small, sedan and SUV cars	Small, sedan and SUV cars	One model of small city cars	Small, sedan and SUV cars	
Walking time to access the vehicle	up to 5 min	up to 5 min	6 to 10 min	up to 5 min	
	○	○	○	○	○

Figure 8: Example of choice task presented to respondents

Table 4 presents all the attributes employed and their corresponding levels. They were selected by considering the published literature and a qualitative survey previously conducted. The levels for the one-time subscription cost attribute and the cost usage attribute in Copenhagen depended on the car-sharing service type, and they are presented in the appendix section A.1. The cost-usage includes many levels to test if different ways of presenting the cost influence respondents' choices.

Attribute	Variable type	Levels of the attributes
Walking time access to the vehicle	Continuous	1: up to 5 min
		2: 6 to 10 min
		3: 11 to 15 min
Probability to get a share vehicle	Continuous	1: 10 out of 10 request
		2: 9 out of 10 request
		3: 7 out of 10 request
Car-sharing vehicle types	Dummy	1: One model of small city cars
		2: Small city cars and sedan cars
		3: Small, sedan and SUV cars
Car-sharing vehicle engine type	Dummy	1: Combustion
		2:Electric
		3:Mix of combustion and electric
Walking time from the parking location to destination	Continuous	1: Up to 5 min
		2: 6 to 10 min
		3: 11 to 15 min
Extra features	Dummy	1: Guaranteed child car seat availability
		2: Family/friends account with discounted rates
		3: A business account with discounted rates
		4: Booking in advance
		5: Plan including other modes for a seamless door to door trip
		6: Collect credits to redeem for goods

Table 4: Description of the attributes of the alternatives

After the experiment, respondents could indicate how frequently they would use the most preferred alternative, and the results were compared with the initially stated frequency. In this way, we can observe whether the selected plan would lead to an increase, decrease, or no change in their car-sharing usage.

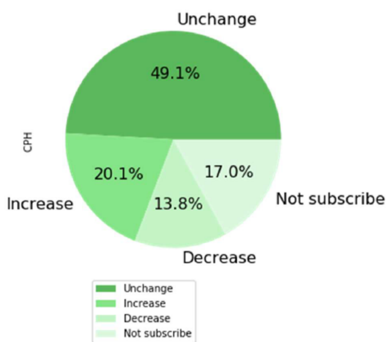

Figure 9: Change in car-sharing usage: Copenhagen's members and past-members

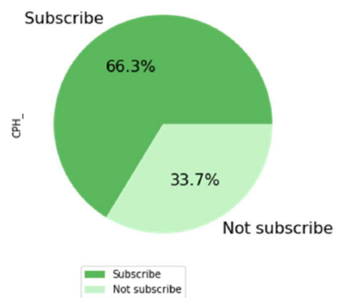

Figure 10: Change in car-sharing usage: Copenhagen's non members

In figure 10, it is interesting to see that 66.3% of the non-members' respondents would become members of car-sharing if their selected plan were available. On the other hand, figure 9 shows that almost 70% of the car-sharing members would increase or remain their use given this new alternative.

5 Results and discussion

In this section we present our results for the previously formulated models, employing the data describe in section 4.

5.1 Definition of the alternatives

In table 5, we present the alternatives of the choice set, that will be equal to all models, with their corresponding abbreviation and number in the choice set.

Alternative number	Alternatives	Abbrevation
1	Roundtrip car-sharing	RT
2	One-way car-sharing - Station based	OWST
3	One-way car-sharing - Free-floating	OWFF
4	Peer-to-peer car-sharing	P2P
5	None of the above	-

Table 5: Alternatives with their corresponding abbreviation and number in the choice set

All alternatives refer to a type of car-sharing.
As the name suggests, one-way car-sharing - Station-based is a car-sharing business type where the cars are in fixed locations around the city. The user can take a vehicle from a car park and return it to the same place or another location fixed by the company. On the contrary, in one-way car-sharing - free-floating, the users can pick or return the car in an available spot in the city without imposing fixed stations. This CS type gives extra freedom to their user in planning their route.
In roundtrip car-sharing (RT), the cars rest in parking garages scattered around the city and the suburbs, where they are picked up and dropped off in the exact location when finished. The driver has no space problems when he wants to return the car and does not overload the available parking slots in the city, as in the free-floating CS. Still, it can be inconvenient to find a nearby parking garage, depending on the destination.
Finally, peer-to-peer (P2P) car-sharing occurs when a vehicle owner rents out its car to other users for a short-time period. Unlike the other CS types, they are individual car owners who offer their vehicles.

5.2 Mixed logit results

We have started our analysis by computing a mixed model without latent classes to find out which attributes and features are more relevant for the individuals when making the decision.
The correspondent utilities are defined by,

$$U_{int} = ASC_i + \beta_{ix} X_{int} + \beta_Z Z_n + \alpha_{in} + \sigma_{CSplans} + \epsilon_{int} \tag{51}$$

where ASC_i is the alternative-specific constant representing the average effect on the utility of all the factors not included in the model. β_{ix} and β_Z are the vectors of coefficients that account for the impact of the attributes of the alternatives X_{int}, and the socioeconomic variables Z_n, on the utility; they are coefficients to be estimated. α_{in} are error terms normally distributed across individuals that capture the correlation among choices for the same individual (panel data), $\sigma_{CSplans}$ accounts fro the correlation between the different alternatives and ϵ_{nit} is the independent and identically distributed error term following an extreme value distribution.

We have estimated the model using PandasBiogeme [34], and the coefficients with their standard deviation are presented in table 6. The final model includes only the variables which coefficients were found stadistically significant.

The cost of usage, subscription and the age have been scaled for their inclusion in the model. Their corresponding scales are,

$$\begin{aligned} X_{it,subscription\ cost\ scaled} &= \frac{X_{it,subscription\ cost}}{100} \\ X_{it,usage\ cost\ scaled} &= X_{it,usage\ cost} * 10 \\ X_{it,age\ scaled} &= \frac{X_{it,age}}{10} \end{aligned} \tag{52}$$

All cost variables were converted to euro with the exchange rate 1 EUR = 7.4434DKK.

On the other hand, $X_{it,hour\_package}$ and $X_{it,day\_package}$ are dummy variables that are activated when the price is presented per hour or per day, respectively, to represent the influence of the price format in the utility.

When different levels of the same categorical variable are included in the model, one is set to zero to reduce the number of model parameters since only the differences in utility matter. A mix of combustion and electric motor is the base level for the CS vehicle engine. (Its beta value is imposed as zero). For the price format, given the cost in minutes is the reference level. The category 'one model of small city car' is the base level for the CS vehicle type attribute. For more information about the attributes' levels presented in the experiment, the reader is referred to table 4 and to the appendix section A.1.

The socio-demographic variables included in the model were age, car-sharing membership, gender, and the presence of kids at home, all included as dummy variables.

Finally, after testing the model with different combinations of the incentives offered (extra features), any of the incentives appeared relevant.

Variable	Estimate	Rob. Std. err
ASC-OWFF	5.75	1.08
ASC-OWST	5.71	1.08
ASC-P2P	7.37	1.07
ASC-RT	5.83	1.09
$\alpha_{panel\ effect} - OWFF$	1.01	0.21
$\alpha_{panel\ effect} - OWST$	1.09	0.22
$\alpha_{panel\ effect} - P2P$	0.91	0.23
$\alpha_{panel\ effect} - RT$	0.70	0.27
$\beta_{One\ time\ subscription\ cost}$	-0.88	0.13
$\beta_{Usage\ cost(OWFF,OWST,RT)}$	-0.07	0.03
$\beta_{Usage\ cost(P2P)}$	-2.79	0.37
$\beta_{Usage\ cost\ per\ day}$	0.76	0.20
$\beta_{Usage\ cost\ per\ hour}$	-0.32	0.18
$\beta_{Only\ combustion\ cars}$	-0.32	0.10
$\sigma_{Only\ electric\ cars}$	0.78	0.22
$\beta_{Probability\ of\ finding\ a\ shared\ car}$	1.03	0.40
$\sigma_{Walking\ time\ access\ vehicle}$	0.06	0.03
$\beta_{Walking\ time\ from\ parking\ to\ destination}$	-0.03	0.01
β_{Age}	-1.23	0.19
$\beta_{CSMember}$	1.15	0.74
$\beta_{Kidsathome}$	1.64	0.73
$\sigma_{CSplans}$	5.00	0.49

Table 6: Mixed logit model results

In table 6 we observe that the ASC corresponding with the P2P alternative shows the highest value, suggesting a preference for this CS service type without considering all other attributes. Since the variables for the panel effect turn out significant, there are some correlations between the choices of the same decision-maker, underlining the presence of unobserved individuals' factors. Moreover, the variance accounting for the corrections between plans shows unobserved common relationships between CS service types.

On the other hand, all cost coefficients are negative, as we expected under the behavioral theory. The usage cost coefficient has the same value for the alternatives OWFF,OWST and RT, but it is significantly different and presents a more negative value for P2P. This coefficient suggests that given specific characteristics of peer-to-peer CS services, users are less willing to pay the same as for the other alternatives, with everything else equal.

Furthermore, if the price format is given per day, it is seen more negatively as given per hour, and the price format per minute is preferred for any CS service. Probably, the cause is that individuals expect to drive CS cars for relatively short trips. The data support this evidence: around 50% of the responders used the shared car for up to one hour.

The coefficient for the walking time to reach the destination is significant but present a small value. However, the walking time to access the car has high heterogeneity. We have modeled

its coefficient with a normal distribution, resulting in a mean value of zero and a standard deviation of 0.06. This result suggests that while some people may find walking to the car uncomfortable, others are more inclined to walk for a few minutes before getting into the car, which is also related to the active mode culture in CPH.
The type of CS engine was also relevant for the analysis. It turned out that responders from CPH see more positively CS services with electric engines rather than a mix of combustion and electric engine (reference level), being the combustion engine type their least preferred. This may be due to concerns about increasing the pollution level, reflected by the Danish environmental culture.
The attribute accounting for the probability of finding a car also relevant and its coefficient is positive since as the probability of finding a car increases, the utilities also have to increase.
For the socio-demographic characteristics, the increase in age negatively affects the likelihood of choosing any CS service. This tendency is also present in [35], where the possible causes were the long-term habit of owning a car and the generation influence.
Being a CS member increases the utilities of choosing any CS service, as we have expected since it seems that the individual have already a predisposition for the service. Having kids at home also increases the probability of choosing a CS service, probably related to the respondent's age that still has their kids at home. Also, mobility literature supports that when there are significant life changes, such as the birth of a child, people become more inclined to use car-sharing [37].
Concerning the income level, although some studies have found a relationship between lower use of CS services among people with a lower income level [36], the income level did not turn out significant for our sample.
The final model log-likelihood has a value of -1975 that can be compared with the log-likelihood of the models presented in the following sections.

5.3 LCCM results

We extended our model once we overviewed the structure and importance of the attributes and socio-demographic variables when no latent variables were present.
In this section, we present the results of the traditional LCCM model without the inclusion of the attitudinal variables into the class membership model. We have only obtained results for the model with two latent classes, as it has not reached convergence when the number of classes has been increased.
The LCCM model has a log-likelihood of -1990.14, the value of the AIC is 4042.27, the BIC is 4209.0 and the number of parameters is 31.
Table 7 shows the corresponding coefficients for both classes.

Variable	Class specific choice model	
	Class 1	Class 2
ASC-OWFF	-3.60(1.30)	2.03(0.33)
ASC-OWST	-3.36(1.29)	2.01(0.34)
ASC-P2P	-0.41(1.37)	3.17(0.37)
ASC-RT	-4.04(1.33)	2.04(0.34)
$\beta_{One\ time\ subscription\ cost}$	-1.00(0.48)	-0.71(0.11)
$\beta_{Usage\ cost(OWFF,OWST,RT)}$	-0.19(0.08)	-0.04(0.03)
$\beta_{Usage\ cost(P2P)}$	-5.81(1.37)	-2.00 (0.30)
$\beta_{Usage\ cost\ per\ day}$	-2.37(0.62)	-0.48(0.17)
$\beta_{Usage\ cost\ per\ hour}$	-1.23(0.45)	-0.12(0.15)
$\beta_{Only\ combustion\ cars}$	-0.17(0.32)	-0.32(0.08)
$\beta_{Probability\ of\ finding\ a\ shared\ car}$	2.32(1.30)	0.73(0.32)
$\beta_{Walking\ time\ from\ parking\ to\ destination}$	0.02(0.04)	-0.03(0.01)

Table 7: Parameters of the LCCM without attitudinal statements

We see a clear difference in the values of the alternative-specific coefficients for both classes. We observe negative values for class one, which imply a negative predisposition to choosing any car-sharing service than any other means of transport, represented by the 'None of the above' alternative. (The representative utility (V_i) for this alternative is the baseline, and therefore it is settled to zero).
In contrast, the alternative-specific values are positive for all the alternatives in class two, representing an inclination for choosing any car-sharing services, where peer-to-peer seems to be the preferred one, without considering the other features.
The cost of subscription and usage coefficients are also more negative for class one than two. Moreover, it is worth mentioning that it is more negative for P2P than for the other alternatives in both classes, which shows that P2P users are more susceptible to price increases than in the case of private use car-sharing. This fact may be related to the predisposition of paying less for a shared service.
For both classes, the price presentation per hour is preferred rather than per day, related to the fact that CS users tend to drive considerably short trips.
On the other hand, for those who are more inclined to use car-sharing services (class 2), a car with a combustion engine influences their decision more negatively than those less prone to the service. The baseline level for the engine type is a combination of electric and combustion engines, which is seen more positively than only combustion cars for both classes.
It is also interesting to mention that the coefficient for the probability of finding a car is higher for class one, which evidences that the availability of the service seems to be more critical for those who are less inclined to use CS.
The walking time for parking to destination is slightly negative for class two and the walking time to access the car-sharing turn out no significant and it has been removed from the model formulation.
Table 8 presents the values of the parameter for the class membership model. Class two is the baseline, and, therefore, all its parameters values are set to zero with no loss of informa-

tion since only the differences in the utility matter.

Variables	**Parameters**
Class-specific constant (Class 1)	-1.28(0.54)
CS member (Class 1)	-0.55(0.30)
Age (Class 1)	0.04(0.01)
Own a car (Class 1)	-0.09(0.26)
Leases a car (Class 1)	-1.18(0.64)
Bike at home (Class 1)	-0.85(0.39)
Kids at home (Class 1)	-0.49(0.29)

Table 8: Parameter of the class membership LCCM model without attitudinal statements

We recall that individuals with a higher probability of belonging to class one are less inclined about the concept of car-sharing, which is consistent with a negative coefficient for CS membership. Higher age is also a predisposition for belonging to class one since there is evidence in the literature [35] suggesting that young people are more prone to use this service. Individuals with bikes or kids at home have less probability of belonging to class one. We have also mentioned that the literature supported kids at home as a possible factor in increasing the likelihood of using CS. Owning a car seems not relevant for this formulation given the high standard deviation of its coefficient compared with its value. Finally, having access to a leased car scores negatively for the probability of class one, probably because these individuals are more used to lease or car companies and therefore, they have more predisposition to CS.

5.4 LCCM with attitudinal statements results

In this section, we present the results of the traditional LCCM model, where the factor scores obtained from a Confirmatory Factor Analysis (CFA) are included in the class membership model as continuous variables.
In this case, the model log-likelihood is -1965.71, the AIC value is 4001.42, the BIC is 4189.0 and the number of parameter is 35. The likelihood value is higher than in the previous formulation, as expected when including more information into the model. However, the AIC and BIC values are smaller than in the last model, which indicates that the inclusion of more parameters don't seem to be overfitting the data more than in the previous formulation.
Table 9 presents the model coefficients. We observed a clear difference in the alternative-specific constants for the two classes.

Variable	**Class specific choice model**	
	Class 1	**Class 2**
ASC-OWFF	-3.35(1.29)	2.00(0.33)
ASC-OWST	-3.16(1.29)	1.98(0.34)
ASC-P2P	-0.17(1.36)	3.12(0.37)
ASC-RT	-3.79(1.32)	2.00(0.34)
$\beta_{One\ time\ subscription\ cost}$	-0.95(0.48)	-0.71(0.11)
$\beta_{Usage\ cost(OWFF,OWST,RT)}$	-0.19(0.09)	-0.05(0.03)
$\beta_{Usage\ cost(P2P)}$	-5.72(1.28)	-2.00 (0.30)
$\beta_{Usage\ cost\ per\ day}$	-2.29(0.63)	-0.50(0.17)
$\beta_{Usage\ cost\ per\ hour}$	-1.17(0.48)	-0.14(0.15)
$\beta_{Only\ combustion\ cars}$	-0.09(0.32)	-0.32(0.08)
$\beta_{Probability\ of\ finding\ a\ shared\ car}$	1.98(1.29)	0.75(0.32)
$\beta_{Walking\ time\ from\ parking\ to\ destination}$	0.01(0.04)	-0.02(0.01)

Table 9: Parameters of the LCCM with attitudinal statements

Class one presents a negative bias about CS services compared with other means of transportation, and class two is more inclined about CS service without accounting for the other features. The values of the betas are similar to the formulation without attitudinal variables. Everything explained in the previous section about the beta coefficients applies as well to this case.
Table 10 presents the values of the parameters for the class membership model. Class two is again the baseline, and all its parameters are set to zero.

Variables	**Parameters**
Class-specific constant (Class 1)	-1.62(0.57)
CS member (Class 1)	-0.20(0.32)
Age (Class 1)	0.05(0.01)
Own a car (Class 1)	-0.59(0.31)
Leases a car (Class 1)	-1.46(0.65)
Bike at home (Class 1)	-0.63(0.40)
Kids at home (Class 1)	-0.40(0.32)
F1 (Class 1)	0.04(0.14)
F2 (Class 1)	-0.11(0.11)
F3 (Class 1)	-0.58(0.15)
F4 (Class 1)	-0.51(0.12)

Table 10: Parameter of the class membership LCCM model with attitudinal statements

Class two represents a bias to CS services in comparison with class one. For this reason, it is not surprising that, as before, older adults have a higher probability of belonging to class one, the one less inclined about CS. As commented before, having kids and bikes at

home decreases the likelihood of belonging to class one. Owning a car and having access to lease cars incline individuals to belong to class two.
Regarding the attitudinal factors, in this case, F1 and F2 do not seem relevant for the analysis since their standard deviation is considerably high compare with its value. This inconvenience will be solved when increasing to three classes in the GBM model since the heterogeneity of the classes will be reduced.
However, a positive score for factors two and three correlates with negative class one likelihood. We recall that factor three corresponds with a positive belief about the CS concept, and factor four represents the agreement that car-sharing is an excellent alternative to car ownership. Therefore, it seems that including these attitudinal indicators improves the characterization of the classes coherently.

5.5 GBM without attitudinal statements results

The model formulation of this section is described in 3.4, where the class membership model is defined as a Gaussian-Bernoulli Mixture Model (GBM).
It is important to note that for the GMM we have four different options to formulate the covariance matrix, depending on the model complexity and the flexibility we want to archive. In the full covariance formulation, each component of the mixture has its general covariance matrix. It allows for the highest flexibility, but it is also the model with more parameters, which can tend to overfit. The tied covariance formulation specifies that all mixture components share the same general covariance matrix, reducing the number of model parameters. On the other hand, no correlation is assumed with the diagonal covariance, where each component has its own diagonal covariance matrix. Finally, the spherical covariance constraints, even more, the covariance matrix assigning to each mixture component one single variance. This model formulation only has one continuous variable in the class membership (age). Therefore, the covariance matrix consists of only one element. We can only compare the model with full covariance (which is equivalent to diagonal and spherical) where the variance is different for the three classes or the tied formulation where the variance is the same for all the mixture components.
We have performed the likelihood ratio test between models with the same number of classes to compare models with different covariance matrices.
On the other hand, we have employed the AIC and BIC goodness-of-fit and its corresponding p-value to select the attributes and socio-demographic variables present in each model.
Table 11 shows the obtained results of the models with 2 and 3 classes and different covariance matrices:

	Classes	Log Likelihood	AIC	BIC	Num of parameters
GMB-LCCM	k=2	-1999.99	4074.64	4273	39
Full covariance	k=3	-1989.63	4097.27	4414	59
GMB-LCCM	k=2	-1999.508	4074.17	4278	38
Tied covariance	**k=3**	**-1966.74**	**4047.48**	**4355**	**57**

Table 11: GBM without attitudinal variables - models specifications

It is clear that the best model specification is the model with 3 classes and tied covariance. The beta values for this model, are presented in table 12, showing a clear difference in the three classes' coefficients.

Variable	**Class specific choice model**		
	Class 1	**Class 2**	**Class 3**
ASC-OWFF	3.74(0.51)	-3.47(1.31)	-1.06(0.76)
ASC-OWST	3.12(0.53)	-3.29(1.30)	0.48(0.71)
ASC-P2P	4.25(0.58)	-0.36(1.40)	1.79(0.73)
ASC-RT	3.08(0.53)	-4.09(1.35)	0.53(0.70)
$\beta_{One\ time\ subscription\ cost}$	-1.04(0.00)	-1.06 (0.48)	-0.43 (0.20)
$\beta_{Usage\ cost(OWFF,OWST,RT)}$	0.06(0.04)	-0.18(0.08)	-0.28(0.07)
$\beta_{Usage\ cost(P2P)}$	-1.23 (0.43)	-5.81 (1.33)	-3.41 (0.66)
$\beta_{Usage\ cost\ per\ day}$	-0.36(0.26)	-2.42(0.64)	-0.80(0.35)
$\beta_{Usage\ cost\ per\ hour}$	-0.11(0.23)	-1.21(0.44)	-0.22(0.31)
$\beta_{Only\ combustion\ cars}$	-0.20(0.13)	-0.07(0.32)	-0.60(0.17)
$\beta_{Probability\ of\ finding\ a\ shared\ car}$	-0.19(0.49)	2.28(1.30)	2.52(0.66)
$\beta_{Walking\ time\ from\ parking\ to\ destination}$	-0.05(0.02)	0.01(0.80)	0.01(0.02)

Table 12: Parameters of the GBM without attitudal variables - class specific choice model

Individuals with a higher probability of belonging to the first latent class are more predisposed to car-sharing without considering the observed features. All the alternative-specific constants have a positive value which means that the mean of all unobserved attributes inclined them to choose an alternative that includes car-sharing compared to any other mean of transport.
On the other hand, latent class two is the least inclined about the concept of car-sharing. The alternative-specific constants of this class are negatives, meaning that non-choosing any car-sharing plan gives them more utility than choosing any CS type without accounting for its characteristics. The cost usage and subscription coefficients are more damaging, evidencing their inclination against car-sharing use.
Class three evidences a behavior between the previously mentioned classes somewhere in the middle. This tendency is also visible in this class beta values, which show a more neutral predisposition towards car-sharing than other means of transport. It is interesting to mention that is the class with less influence in the subscription cost, but more negatively affected by the cost usage of all the alternatives, except P2P. Moreover, it is also the more concerned if the type of engine car is combustion.
P2P presents a higher alternative-specific value than the other alternatives and for all the classes. And table 12 also shows a bias towards displaying the price per minute instead of per day or hour across all the classes, related to the fact that car-sharing users tend to drive it for short periods.
It is also worth mentioning that the beta corresponding with the cost usage for all the alternatives except P2P for the first latent class has a positive value, which contradicts the

theory of choice models, even if this coefficient has a large standard deviation. This issue will be solved in the subsequent models when including attitudinal variables in the class membership.
If the engine type of the car is combustion, it scores more negatively for the classes that are more inclined about the concept of car-sharing than for the least willing. Therefore, car-sharing could be seen as an electric alternative to those who are more worried about the environmental footprint of their trips.
Regarding the probability of finding a car, it is a more important feature for those less inclined to use car-sharing and those more neutrally inclined, which are more dependent on the availability of the service.
The distribution of the socio-demographic variables in the class membership model is presented in table 13. The continuous variable age is standardized, which means that a latent class with a negative value of age is characterized by individuals younger than the average. For the categorical variables, the values shown in table 13 are the probability of presenting the feature or not, given that the individual belongs to that class.

Variables		**Class 1**	**Class 2**	**Class 3**
Age	Continuous	-0.324	0.464	-0.234
Own a car	Yes	0.652	0.755	0.633
	No	0.348	0.245	0.367
Leases cars	Yes	0.082	0.029	0.064
	No	0.918	0.971	0.936
Bike at home	Yes	0.974	0.869	0.944
	No	0.026	0.131	0.056
Kids at home	Yes	0.332	0.116	0.189
	No	0.668	0.884	0.811
CS Member	Yes	0.227	0.113	0.195
	No	0.773	0.887	0.805

Table 13: Mean matrix of the class membership model (GBM without attitudinal variables)-Tied covariance K=3

Individuals with a higher probability of belonging to class one, the most inclined to use car-sharing services, tend to be the younger ones with more kids at home and more access to a leased or company car. They are also the most likely to be CS members and have at least one bike at home compared to the other classes.
However, individuals with a higher probability for class two, the least inclined about CS, tend to be older than the average. In addition, they owned more cars and fewer bikes than individuals with a predisposition for the other classes, and they are the ones with fewer kids at home. They tend to be less likely to be car-sharing members, which correlates with the results of its class choice model, and almost no one has access to a leased or company car.
While class two grouped individuals with opposite socio-demographic characteristics than class one. Class three is a trade-off between them for all the socio-demographic characteristics. For example, individuals with a higher probability of belonging to this class tend to be

younger than the average but still a bit older than individuals with a predisposition to class one.
To have a clear idea of the class configuration, we plot the percentage of the socio-demographic variables and the average value for the age for each class in figure 11.

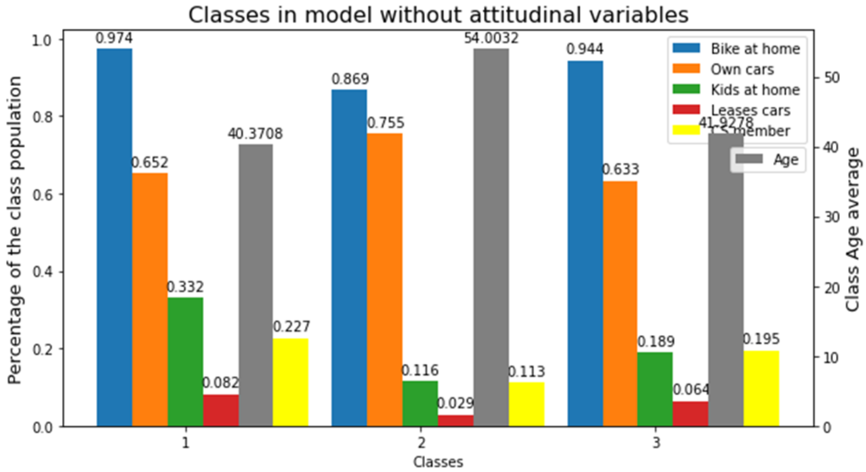

Figure 11: Classes configuration for GBM model without attitudinal variables

5.6 GBM with attitudinal statements results

We go one step further in the analysis by incorporating attitudinal variables into the model, which we expect will give us more insight into the decision behavior. We included the information of the attitudinal statements as psychological indicators in the class membership model. Table 14, shows the specifications for models with a different number of classes and types of covariance matrices. As in the previous section, the likelihood ratio test has been employed to select different covariance matrices within the models with the same number of classes.

	Classes	Log Likelihood	AIC	BIC	Num of parameters
GMB-LCCM	K=2	-1980.84	4111.7	4515	75
Full covariance	K=3	-2006.72	4239.44	4847.0	113
GMB-LCCM	K=2	-1980.64	4081.3	4404	60
Tied covariance	**K=3**	**-1966.11**	**4098.2**	**4544**	**83**

Table 14: GMB with attitudinal variables - models specification

The models are sensitive to starting values and could not guarantee convergence to the global maximum. We have employed different sets of random initialization points for the models with two classes, all of which have reached convergence.
On the other hand, it was not easy for the models with three classes to arrive at a global maximum. We have employed as initial values the coefficients from the previous models

with two classes, and we have set the parameters of the third class to zero. The solution was stable with this initialization but did not guarantee to find the global maximum. Thus, we have employed random initialization, in the case of the model with tied covariance and three classes, which has allowed us to reach a better value of likelihood.
Therefore, the convergence of the models needs to be analyzed further. However, we think that including more data would have been the first step to having more stable models since the number of parameters is high compared with the number of observations. The number of observations used to estimate each parameter could become insufficient in some cases.
We have also tried to model four classes, but we could not find a model that converged, probably due again to the small size of the dataset compared with the number of parameters.
The models with tied covariance have been selected. They have a higher likelihood value, employing fewer parameters, which also prevents overfitting.
In table 15, we present the parameters of the model with three classes and tied covariance.

Variable	**Class specific choice model**		
	Class 1	**Class 2**	**Class 3**
ASC-OWFF	-3.60(1.32)	-3.64(2.07)	2.36(0.36)
ASC-OWST	-3.47(1.32)	-1.62(1.75)	2.31(0.37)
ASC-P2P	-0.54(1.42	2.04(1.77)	3.21(0.40)
ASC-RT	-4.2365(1.40)	-0.57(1.74)	2.19(0.37)
$\beta_{One\ time\ subscription\ cost}$	-0.80(0.47)	-0.27 (0.49)	-0.61(0.00)
$\beta_{Usage\ cost(OWFF,OWST,RT)}$	-0.18(0.04)	-0.18(0.08)	-0.02(0.03)
$\beta_{Usage\ cost(P2P)}$	-5.66(1.26)	-6.5(1.5)	-1.66(0.32)
$\beta_{Usage\ cost\ per\ day}$	-2.30(0.66)	-2.34(0.85)	-0.40(0.18)
$\beta_{Usage\ cost\ per\ hour}$	-1.09(0.45)	-1.26(0.71)	-0.07(0.16)
$\beta_{Only\ combustion\ cars}$	0.03(0.32)	-2.06(0.54)	-0.22(0.09)
$\beta_{Probability\ of\ finding\ a\ shared\ car}$	2.23(1.32)	6.14(1.74)	0.42(0.34)
$\beta_{Walking\ time\ from\ parking\ to\ destination}$	0.00(0.03)	0.00(0.05)	-0.03(0.01)

Table 15: Parameters of the GBM with attitudal variables (K=3, tied covariance) - class specific choice model

In this case, we can see the same tendency as before, where class one is not inclined about the concept of car-sharing (negative alternative-specific coefficients), while class three is predisposed to car-sharing (positive alternative-specific coefficients), regardless of its characteristics.
Moreover, we also observed a class with a more neutral predisposition about car-sharing, as before. Even some alternative-specific have a negative sign, the standard deviation is high, evidencing a significant heterogeneity around zero for these coefficients.
The P2P car-sharing alternative continues to be the preferred one across all the classes, even an increase in its price is viewed more negatively, as we have commented before. The price presentation per hour instead of per day is more convenient for all classes, even though the price per minute is still preferable.
Cars with combustion engines are seen unfavorable in classes three and two, specially in this

last one; while it is not significant in class one, the less inclined about CS services.
On the other hand, class two, the more neutrally incline about the CS concept, also presents a high value for the probability of finding a car-sharing, which shows a potential increase in CS use for individuals in this class if CS services maintain high rates of availability.
Finally, the walking time to arrive at the destination seems to be a minor drawback for decision-makers with a predisposition to CS usage. However, its value is not so crucial in the decision for individuals in class one and two.
Traditionally, the estimated cost and time coefficients ratios provide information on the value of time. By definition, the value of time is the extra cost a person would be willing to incur to save time. However, our application does not include the time parameter as a coefficient into the model and this value can not be compared for the different classes.
Finally, table 16, we present the information about the socio-demographic variables together with the factors extracted from the CFA of the attitudinal questions that provide us extra insights into the class membership model.

Variables		**Class 1**	**Class 2**	**Class 3**
Age	Continuous	0.474	-0.543	-0.256
F1	Continuous	-0.165	0.921	-0.008
F2	Continuous	0.071	-1.178	0.105
F3	Continuous	-0.294	0.173	0.180
F4	Continuous	-0.386	0.504	0.200
Own a car	Yes	0.765	0.159	0.702
	No	0.244	0.841	0.298
Leases cars	Yes	0.029	0.088	0.073
	No	0.971	0.912	0.927
Bike at home	Yes	0.867	1.000	0.957
	No	0.133	0.000	0.043
Kids at home	Yes	0.116	0.000	0.306
	No	0.884	1.000	0.694
CS Member	Yes	0.114	0.195	0.215
	No	0.886	0.805	0.785

Table 16: Mean matrix of the class membership model (GBM with attitudinal variables)-Tied covariance K=3

In this case, the socio-demographic distributions in class one is similar to the GBM model without factors and three classes. People with a higher probability of belonging to this class tend to have more cars and fewer bikes at home than the other classes. Also, it has the lowest percentage of CS members and people with access to lease cars of the three classes.
However, it is interesting to mention that the configuration of class two changes with the inclusion of the attitudinal factors. Individuals with a higher probability of belonging to class two tend to be very young people, without any kids at home, with bikes, and it is

the class with the highest likelihood of having access to lease or company cars and the least incline to own a car. In addition, the probability of being a CS member is around 20%, a bit above the average, considering that the sample percentage is 17.5% for the individuals living in CPH.
On the other hand, class three is also represented by individuals younger than the average. It is the class with the highest probability of being a CS member and having kids at home. According to the literature, these facts support car-sharing services, as we have mentioned before. The other features present values in the middle between the other classes means.
In figure 12, we represent the distribution of the socio-demographic variables more visually.

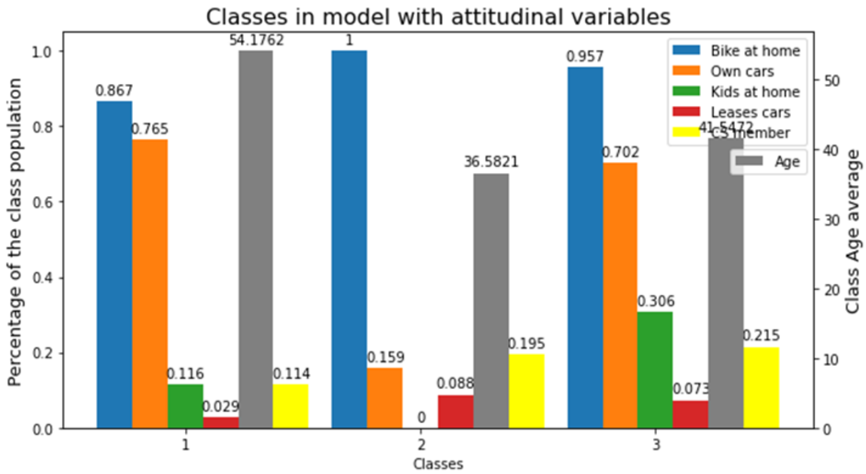

Figure 12: Classes configuration for GBM with attitudinal variables

Regarding the attitudinal indicators included through the scores factor, we see that class one has the fewest car-related issues, which is consistent with being the class with more percentage of car ownership. The factor accounting of the inclination toward the car-sharing concept is below the average, as well as the belief that CS is a convenient alternative to car ownership, as we expected.
On the other hand, people belonging to class two have more car-related issues. This fact directly connects with being the class where people own fewer cars, only around the 16%. They strongly disagree with the affirmation that driving a private vehicle is the most convenient way to move around, and they are inclined to think positively about car-sharing. They believe CS is an excellent alternative to car ownership, and they wouldn't mind sharing their car with other people.
Class three is the most inclined to think positively about car-sharing services according to its value for factor three, which is also reflected in its alternative-specific coefficients. They also agree that CS is an excellent alternative to car ownership above the average of sample opinions. Surprisingly, people in this class tend to agree that driving a private car is the most convenient way to move around.
Figure 13 show us a more visual distribution of the standardized factor scores for each class.

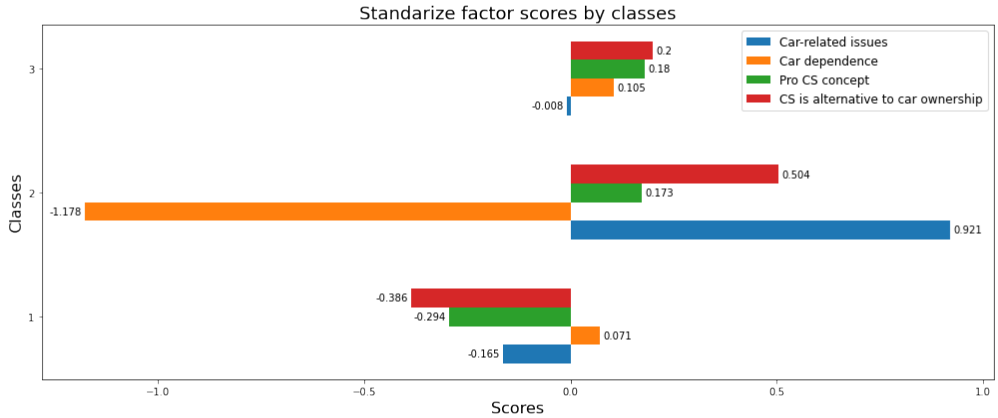

Figure 13: Standardized factor scores for each class - GBM model

5.7 Comparison model with & without attitudinal statements

In this section, we compare the GBM model with three classes, including the attitudinal indicators and without them, and how individuals in the sample change class with the inclusion of attitudinal information in the model.
Our comparison is limited since the GBM is a probabilistic approach model, where each individual has an associated probability of belonging to each class. For the visualization of figure 14 we have assigned to each individual the class with the highest probability of belonging, which produces a bias and underestimates the classes with low shares. It is also important to note that the classes without attitudinal statements are not the same as with them since they are two different models and the configuration of socio-demographic variables is different.
However, we consider interesting to visualize how most individuals with the highest probability of belonging to the class more inclined about the CS concept (according to its alternative-specific values) are also more inclined about CS when we include the attitudinal statement. The opposite applies to the individuals belonging to the class least prone to CS.
The second insight that we can get from figure 14 is that the probabilities of belonging to each category in the model with attitudinal statements are more extreme. Because it assigns individuals more equitably to the different classes when the assignment criterion is the class with the highest probability. This can help us design more targeted policies since we can better characterize a population group when their class assignment is more defined.

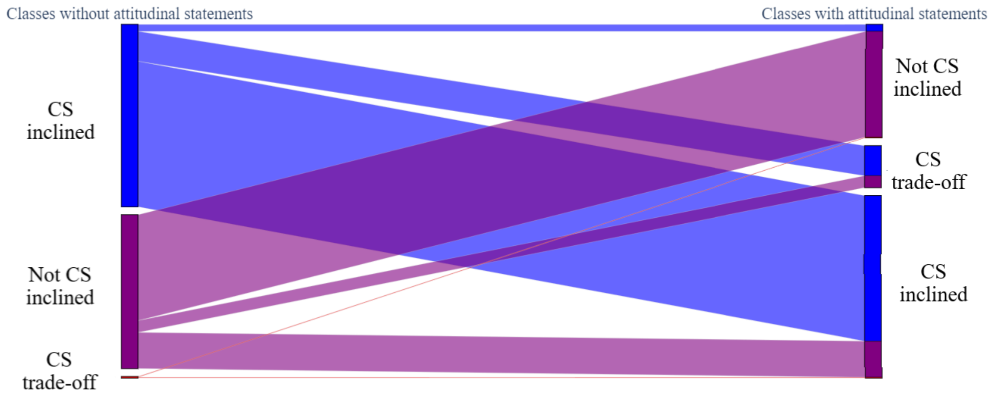

Figure 14: Parallel category plot - classes in models with and without attitudinal information

5.8 Comparison LCCM & GBM model results

In this section, we will compare the models with attitudinal question and two classes for the LCCM and the GBM. For the LCCM approach no more than two classes could be estimated due to convergence problems.
Figure 15 shows that the parameters estimates of both class-specific choice models are almost the same.

Variable	**Class specific choice model**	
	Class 1	**Class 2**
ASC-OWFF	-3.33(1.29)	1.97(0.33)
ASC-OWST	-3.13(1.32)	1.95(0.34)
ASC-P2P	-0.14(1.37)	3.10(0.37)
ASC-RT	-3.84(1.32)	1.97(0.34)
$\beta_{One\ time\ subscription\ cost}$	-0.86(0.47)	-0.71 (0.11)
$\beta_{Usage\ cost(OWFF,OWST,RT)}$	-0.20(0.08)	-0.05(0.03)
$\beta_{Usage\ cost(P2P)}$	-5.60 (1.25)	-2.00 (0.29)
$\beta_{Usage\ cost\ per\ day}$	-2.39(0.65)	-0.49(0.17)
$\beta_{Usage\ cost\ per\ hour}$	-1.18(0.45)	-0.13(0.15)
$\beta_{Only\ combustion\ cars}$	-0.05(0.32)	-0.32(0.08)
$\beta_{Probability\ of\ finding\ a\ shared\ car}$	2.00(1.28)	0.75(0.32)
$\beta_{Walking\ time\ from\ parking\ to\ destination}$	0.01(0.04)	-0.02(0.01)

Table 17: Parameters of the GBM with attitudal (K=2, tied covariance)

Variable	**Class specific choice model**	
	Class 1	**Class 2**
ASC-OWFF	-3.35(1.29)	2.00(0.33)
ASC-OWST	-3.16(1.29)	1.98(0.34)
ASC-P2P	-0.17(1.36)	3.12(0.37)
ASC-RT	-3.79(1.32)	2.00(0.34)
$\beta_{One\ time\ subscription\ cost}$	-0.95(0.48)	-0.71(0.11)
$\beta_{Usage\ cost(OWFF,OWST,RT)}$	-0.19(0.09)	-0.05(0.03)
$\beta_{Usage\ cost(P2P)}$	-5.72(1.28)	-2.00 (0.30)
$\beta_{Usage\ cost\ per\ day}$	-2.29(0.63)	-0.50(0.17)
$\beta_{Usage\ cost\ per\ hour}$	-1.17(0.48)	-0.14(0.15)
$\beta_{Only\ combustion\ cars}$	-0.09(0.32)	-0.32(0.08)
$\beta_{Probability\ of\ finding\ a\ shared\ car}$	1.98(1.29)	0.75(0.32)
$\beta_{Walking\ time\ from\ parking\ to\ destination}$	0.01(0.04)	-0.02(0.01)

Table 18: Parameters of the LCCM with attitudinal statements

Figure 15: Comparison LCCM and GBM-LCCM choice models (K=2

Therefore the changes in the model specifications are cause by changes in the membership models. They are presented together in figure 16.

Variables		Class 1	Class 2
Age	Continuous	0.466	-0.289
F1	Continuous	-0.148	0.091
F2	Continuous	-0.052	-0.032
F3	Continuous	-0.296	0.183
F4	Continuous	-0.379	0.235
Own a car	Yes	0.756	0.643
	No	0.244	0.357
Leases cars	Yes	0.029	0.075
	No	0.971	0.925
Bike at home	Yes	0.868	0.962
	No	0.132	0.038
Kids at home	Yes	0.115	0.272
	No	0.885	0.728
CS Member	Yes	0.112	0.214
	No	0.888	0.786

Mean matrix of the class membership model (GBM with attitudinal variables)

Variables	Parameters
Class-specific constant (Class 1)	-1.62(0.57)
CS member (Class 1)	-0.20(0.32)
Age (Class 1)	0.05(0.01)
Own a car (Class 1)	-0.59(0.31)
Leases a car (Class 1)	-1.46(0.65)
Bike at home (Class 1)	-0.63(0.40)
Kids at home (Class 1)	-0.40(0.32)
F1 (Class 1)	0.04(0.14)
F2 (Class 1)	-0.11(0.11)
F3 (Class 1)	-0.58(0.15)
F4 (Class 1)	-0.51(0.12)

Parameter of the class membership LCCM model with attitudinal statements

Figure 16: Comparison LCCM and GBM-LCCM class membership models (K=2)

Even if the likelihood value is better for the LCCM model in this case, we can conclude that the GMB-LCCM allows us to capture more heterogeneity than the LCCM since it can capture up to three latent classes. The GBM has more flexibility than the linear-in-parameters utility specification of the latent classes of the LCCM.
In addition, the specification of the classes becomes more easily interpretable in the GBM model since it returns us the corresponding percentage for each attribute in each of the classes.

5.9 Embeddings model results

We had trained and tested our data both with the E-MNL and the EL-MNL formulation, explained in section 3.6.1 and 3.6.2, respectively.
We have chosen the E-MNL formulation for the interpretation of the embedding space. Even though this model has a lower log-likelihood than the EL-MNL model in the train set, the likelihood increases in the test set. This indicates that the EL-MNL model, which is more complex, may overfit the data.
The analysis of the beta values is not comparable with the other models since no latent classes are present. Therefore, we based our results on the interpretability of the embeddings' dimensions, which provides more insights into the influence of the attitudinal statements on the choice.
We present hereafter the value of the embeddings scaled by the betas, allowing us to compare all the categorical variables together. A higher positive value along each alternative-specific axis indicates a higher positive effect on the corresponding alternative. The opposite happens if the value is negative.

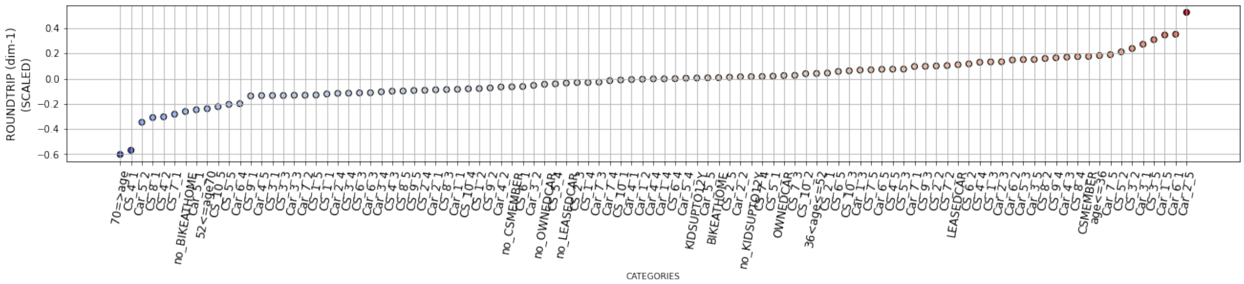

Figure 17: Dimesion 1-(RT) of embeddings space

For dimension one that maps with alternative 1 (Roundtrip-CS), the highest values of the embeddings, which can be interpreted as a tendency for choosing this alternative, correspond with totally agree with the statement that it is difficult to find parking and totally disagree that a car is a status symbol, followed by totally agreeing with the feeling of stress when driving and that car-sharing prevent to not deal with vehicle maintenance and repair, in this order.
The lowest two values of this embedding dimension, in other words, the categorical variables that minor increase the utility of choosing this alternative, everything else being equal, correspond with totally disagree this the previous statement and being more than 70 years.

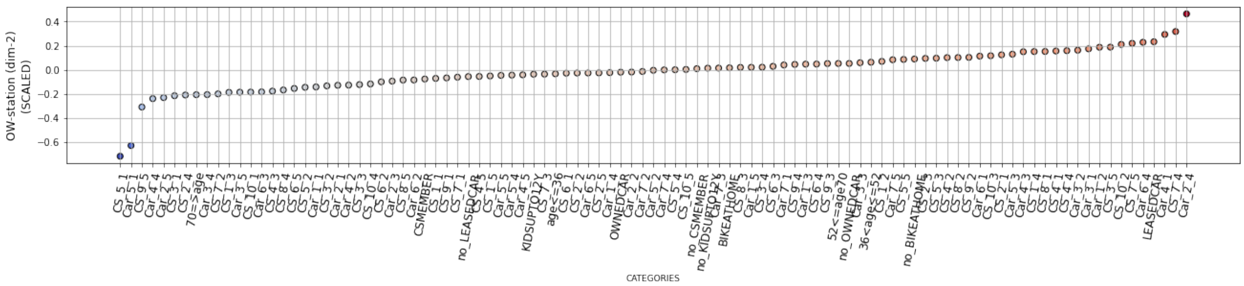

Figure 18: Dimesion 2-(OWST) of embeddings space

On the other hand, in the second dimension of the embedding space (OWST car-sharing), the highest values of the embeddings correspond with the categories of considering challenging to find parking and thinking car-sharing makes life easier. Together with the total disagreement that it is easy to conduct daily trips without a private car and to have access to a leased or company car.
The lowest values in this dimension correspond with a total disagreement that car-sharing help to save from the fuel, taxes, insurance, and parking expenses, together with a total disagreement that the car is a considerable expense.

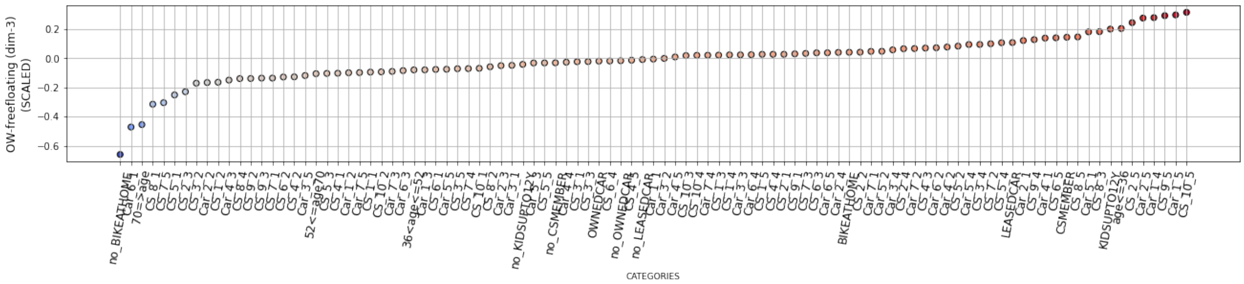

Figure 19: Dimesion 3-(OWFF) of embeddings space

For the third dimension of the embedding space connected with the alternative OWFF car-sharing, the highest values correspond with not minding sharing their car with other people, feeling stress when driving, and thinking that car-sharing is more convenient than public transport. A total agreement with the difficultly of finding parking and that car-sharing is a more environmentally friendly alternative to car ownership. The lowest values, which everything else being equal, reduce the probability of choosing this type of CS, correspond with not having a bike at home, being more than 70 years old.

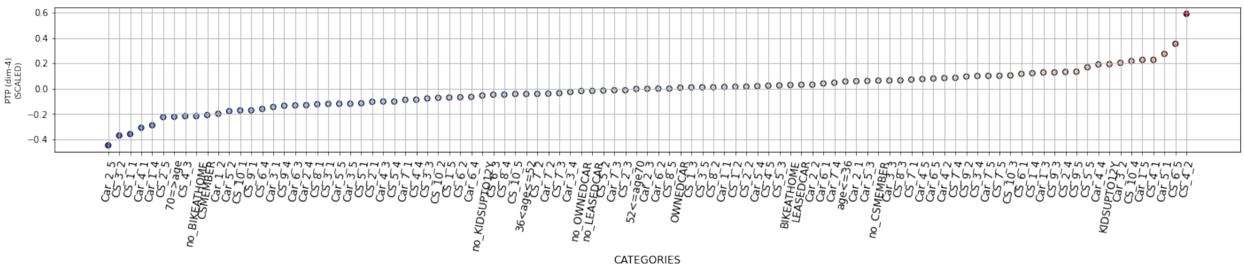

Figure 20: Dimesion 4-(P2P) of embeddings space

For the fourth dimension of car sharing, it is interesting that the highest value correspond with a disagreement about the possibility of CS for saving from the fuel, taxes, insurance, and parking expenses associated with private vehicle ownership.
They totally agree that using a car sharing service can lower my transport expenses and totally disagree Driving a private car is the most convenient way to move around, a total agreement that with the feeling of stress when driving, and an agreement to share their personal car with other people.
The lowest two values correspond with believing that the car is also a status symbol and disagreeing about the difficulty of finding parking.

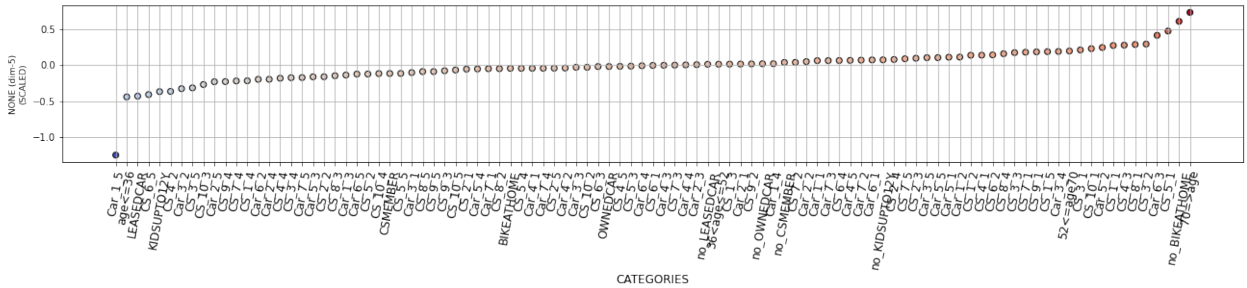

Figure 21: Dimension 5-(None of the above) embeddings space

The last dimension of the embeddings space corresponds with not selecting any of the car-sharing services offered in the other alternatives. The categories with the highest values in this dimension are being over 70 years old and not having a bike at home. Also, disagree that having access to different vehicles is an advantage and a neutral belief about the car as a status symbol. Also, the disagreement with do not have to deal with vehicle maintenance and repair when using CS. The lowest two values of this embedding dimension correspond with totally agree with feeling stress when driving and being less than 36 years old.
We have observed the tendency that young people are more likely to subscribe to a CS plan on several embedding dimensions.

5.10 Discussion

All benchmark models have been evaluated according to goodness-of-fit measures, interpretation of the latent classes and parameter estimates signs of the class-specific DCMs. Likelihood ratio tests have been used for selecting between different types of covariance matrices for the GBM-LCCM and to detect possible over-fitting.
Our results suggest that the inclusion of attitudinal variables not only improves the estimation, which is expected when including more information, but also modifies the configuration of the classes. In addiction, the DCM becomes more behaviorally realistic. For example, individuals who are inclined towards the concept of CS tend to be grouped together in clusters with higher parameter estimates in the utility of choosing CS plans. This indicates that beliefs and attitudes play a key role in decision-making and including this information allows for a more accurate estimation and a better understanding of the classes.
The LCCM has better goodness-of-fit measures for the models with and without attitudinal indicators. However, it was only able to identify two classes, while the GBM-LCCM can identify up to three. Moreover, the specification of the classes for the GBM model is more comprehensive than the LCCM because it can provide the mean of the attitudes scores and socio-demographic characteristics for each group. In addition, it is more flexible since it is not restricted to linear equations for the formulation of the classes, as in the case of the LCCM.

6 Conclusions

We have compared several discrete choice model formulations. We started with a model without latent classes to identify the more relevant features, and we have gradually increased the complexity of the models.

We focused on the relevance and impact of the inclusion of attitudinal information in the models, evidencing its importance in the configuration of the latent classes. It allows us for extra insights to help us design better policies by having a more realistic population segmentation. When creating new approaches to attract and retain car-sharing members, attitudinal information allows us to focus on changing beliefs and mentalities that we know are closely related to individual choices. Finally, it also helps to understand better how the decision-making process works from a theoretical point of view.

Regarding our car-sharing application, we have also improved our understanding of how the features of the car-sharing business can maintain and attract new members.

We have noted the importance of the subscription and usage cost throughout all the models. In addition, the price format is relevant for the decision, and the price per minute is the most attractive. We have also observed that the availability of cars is essential in attracting new members, which can be seen as an improvement opportunity for car-sharing companies. Choosing an electric car is more favorable for those more inclined to use car-sharing, so having a fleet of electric vehicles can be essential to maintain CS members. Finally, while the time to reach the car did not appear relevant in Copenhagen, some efforts could be made to reduce the walking time from the parking until the destination, which negatively influences the choice of a CS service. One solution to this problem could be providing more parking slots.

6.1 Limitations

We have not had time to incorporate latent classes into the embeddings models. Although we designed a simple ANN model to include latent classes, the configuration of the model and the method employed to estimate the ANN, called backpropagation, did not allow us to create a model that integrates the embedding information and updates the class membership simultaneously. Without a doubt, this is a clear line of research that may bear fruit in the future.

The second major limitation is that for the GBM-LCCM model, we have not carried out a joint estimation of the choice model, the class membership, and the attitudinal questions together. Instead, we have included a pre-compute factor analysis into the class membership without considering the respondent's choice when making the configuration of the factors.

A model where psychometric indicators are included in the GBM using Structural Equation Model (SEM) can improve the representation of the latent classes.

The data used for the analysis is not very large, only about 520 individuals, which makes the model estimations complicated when the number of parameters increases so much that there are no sufficient observations to estimate each coefficient. The inclusion of data from Tel-Aviv and Munich, obtained in the same survey, would improve this limitation. Furthermore, it is worth mentioning that the data is not balanced regarding CS membership, as only 17.5% of the Copenhagen respondents are CS members.

Although all the limitations presented in this work, we are optimistic that this analysis has opened the door to future research on integrating attitudinal variables with class membership models through machine learning techniques. This investigation could improve the overall model fit and prediction accuracy, thanks to the discrete representation of heterogeneity due to machine learning techniques' ability to capture complex unobserved patterns. However, we have always to bear in mind that the transparency and interpretability of the models are of utmost importance. A model with high accuracy is of no use if we cannot understand it and draw conclusions since we aim to design better policies and, in conclusion, have a better understanding of the individual's decision processes.

List of Figures

List of Tables

Nomenclature

AIC Akaike information criterion

BMM Bernoulli Mixture Model

CFA Confimatory Factor Analysis

CPH Copenhagen

CS Car-sharing

DCM Discrete Choice Model

EFA Explanatory Factor Analysis

GBM Gaussian-Bernoulli Mixture Mode

GBM-LCCM Gaussian-Bernoulli Mixture Latent Class Choice Model

GMM Gaussian Mixture Model

IIA Independence of Irrelevant Alternatives

KMO Kaiser-Meyer-Olkin

LCCM Latent Class Choice Model

NLP Natural Language Processing

OWFF One-way car-sharing - Free-floating

OWST One-way car-sharing - Station based

P2P Peer-to-peer

RT Roundtrip car-sharing

RUM Random Utility Model

SP State Preference

Bibliography

[1] Frenkel, A., Shiftan, Y., Gal-Tzur, A., Tavory, S. S., Lerner, O., Antoniou, C., Cantelmo, G., Amini, R. E., Lima Azevedo, C. M., Monteiro, M.M., Kamargianni, M., Shachar, F. S., Israel, D., Behrisch, C., Schiff, K., Shalev, J., Peretz, D.(2021). Share More: Shared MObility Rewards - Summary report. https://orbit.dtu.dk/en/publications/share-more-shared-mobility-rewards-summary-report

[2] Hurtubia, R., Nguyen, M. H., Glerum, A., & Bierlaire, M. (2014). Integrating psychometric indicators in latent class choice models. Transportation Research Part A: Policy and Practice, 64, 135-146.

[3] Shaheen, S., A. Cohen, and M. Jaffee. Innovative Mobility: Carsharing Outlook - Worldwide Carsharing Growth. 2020.

[4] Clewlow, R. R. Carsharing and Sustainable Travel Behavior: Results from the San Francisco Bay Area. *Transport Policy*, Vol. 51, No. 2016, 2016, pp. 158-164.

[5] Martin, E., and S. Shaheen. The Impact of Carsharing on Public Transit and Non-Motorized Travel: An Exploration of North American Carsharing Survey Data. Energies, Vol. 4, No. 11, 2011, pp. 2094â2114. https://doi.org/10.3390/en4112094.

[6] Martin, E. W., and S. A. Shaheen. Greenhouse Gas Emission Impacts of Carsharing in North America. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 12, No. 4, 2011, pp. 1074-1086. https://doi.org/10.1109/TITS.2011.2158539.

[7] Chen, T. D., and K. M. Kockelman. Carsharing's Life-Cycle Impacts on Energy Use and Greenhouse Gas Emissions. *Transportation Research Part D: Transport and Environment*, Vol. 47, 2016, pp. 276-284. https://doi.org/10.1016/j.trd.2016.05.012.

[8] Alisoltani, N., L. Leclercq, and M. Zargayouna. Can Dynamic Ride-Sharing Reduce Traffic Congestion? *Transportation Research Part B: Methodological*, Vol. 145, 2021, pp. 212-246. https://doi.org/10.1016/j.trb.2021.01.004.

[9] Millard-Bal, A., G. Murray, J. Ter Schure, C. Fox, and J. Burkhardt.Car-Sharing: Where and How It Succeeds. 2005.

[10] Efthymiou, D., and C. Antoniou. Modeling the Propensity to Join Carsharing Using Hybrid Choice Models and Mixed Survey Data. Transport Policy, Vol. 51, 2016, pp. 143â149. https://doi.org/10.1016/j.tranpol.2016.07.001.

[11] Becker, H., F. Ciari, and K. W. Axhausen. Comparing Car-Sharing Schemes in Switzerland: User Groups and Usage Patterns. Transportation Research Part A: Policy and Practice, Vol. 97, 2017, pp. 17â29. https://doi.org/10.1016/j.tra.2017.01.004.

[12] Monteiro, M.M., Lima Azevedo, C.M., Kamargianni, M., Shiftan, Y., Gal-Tzur, A., Tavory, S., Antoniou, C., Cantelmo, G.. Car-Sharing Subscription Preferences: The

Case of Copenhagen, Munich, and Tel Aviv-Yafo. To be Presented at the 101st Annual Meeting of Transportation Research Board. 9-13 January 2022.

[13] McFadden, D., Train, K. (2000). Mixed MNL models for discrete response. Journal of applied econometrics 15: 447-470.

[14] Walker, J., & Ben-Akiva, M. (2002). Generalized random utility model. Mathematical social sciences, 43(3), 303-343.

[15] Gopinath, D.A., 1995. Modeling Heterogeneity in Discrete Choice Processes: Application to Travel Demand. Ph.D. Dissertation, Massachusetts Institute of Technology.

[16] Molin, E., Mokhtarian, P., Kroesen, M., 2016. Multimodal travel groups and attitudes: a latent class cluster analysis of Dutch travelers. Transp. Res. Part A Policy Pract. 83, 14−29. https://doi.org/10.1016/j.tra.2015.11.001.

[17] Haustein, S., Hunecke, M., 2013. Identifying target groups for environmentally sustainable transport: assessment of different segmentation approaches. Curr. Opin. Environ. Sustain. https://doi.org/10.1016/j.cosust.2013.04.009.

[18] Haustein, S., 2012. Mobility behavior of the elderly: An attitude-based segmentation approach for a heterogeneous target group. Transportation (Amst) 39, 1079â1103. https://doi.org/10.1007/s11116-011-9380-7

[19] Likert, R. (1932) A technique for the measurement of attitudes. Arch. Psychol. 22 (140).

[20] Atasoy, B., Glerum, A., & Bierlaire, M. (2011). Mode choice attitudinal latent class: a Swiss case-study. Second International Choice Modeling Conference, July 2011

[21] Krueger, R., Vij, A., Rashidi, T.H..(2016). Normative belief and modality styles: a latent class and latent variable model of travel behaviour. Springer Science+Business Media New York.

[22] Hurtubia, R., Nguyen, M. H., Glerum, A., & Bierlaire, M. (2020). Drivers and barriers in adopting Mobility as Service (MaaS)- A latent class cluster analysis of attitudes. Transportation Research Part A 123, 378-401.

[23] Train K.,(2009) *Discrete Choice Methods with Simulation*, Cambridge University Press. 1-8.

[24] Train K.,(2009) *Discrete Choice Methods with Simulation*, Cambridge University Press. 41-55.

[25] Thurstone, L. L. (1927). A law of comparative judgment. Psychological Review, 34(4), 273â286. https://doi.org/10.1037/h0070288

[26] Marschak, J.(1960) Binary-Choice Constraints and Random Utility Indicators. Economic Information, Decision, and Prediction pp 218-239

[27] Arkoudi, I., Azevedo, C. L., Pereira, F. C. (2021). Combining Discrete Choice Models and Neural Networks through Embeddings: Formulation, Interpretability and Performance. arXiv preprint arXiv:2109.12042.

[28] Pereira, F. C. (2019). Rethinking travel behavior modeling representations through embeddings. arXiv preprint arXiv:1909.00154.

[29] Sifringer, B., Lurkin, V., Alahi, A. (2020). Enhancing discrete choice models with representation learning. Transportation Research Part B: Methodological, 140, 236-261.

[30] Verwimp, L., Pelemans, J., Wambacq, P. (2015). Expanding n-gram training data for language models based on morpho-syntactic transformations. Computational Linguistics in the Netherlands Journal, 5, 49-64.

[31] Sfeir, G., Abou-Zeid, M., Rodrigues, F., Pereira, F.C., Kaysi, I. (2021). Latent class choice model with a flexible class membership component: A mixture model approach. Journal of Choice Modelling, 41, 100320.

[32] Biernacki, C., Celeux, G., Govaert, G., 2000. Assessing a Mixture Model for Clustering with Integrated Completed likelihood. IEEE Trans. Pattern Anal. Mach. Intell. 22, 19â25. https://doi.org/10.1109/TPAMI.2012.125

[33] Field, A., (2009). Discovering Statistics Using IBM SPSS Statistics, third ed. SAGE Publications Ltd, London.

[34] Bierlaire, M.(2020) A Short Introduction to PandasBiogeme.

[35] Prieto, M., G. Baltas, and V. Stan. Car Sharing Adoption Intention in Urban Areas: What Are the Key Sociodemographic Drivers? Transportation Research Part A: Policy and Practice, Vol. 101, 2017, pp. 218â227. https://doi.org/10.1016/j.tra.2017.05.012.

[36] Efthymiou, D., and C. Antoniou. Modeling the Propensity to Join Carsharing Using Hybrid Choice Models and Mixed Survey Data. Transport Policy, Vol. 51, 2016, pp. 143â149. https://doi.org/10.1016/j.tranpol.2016.07.001.

[37] Priya Uteng, T., T. E. Julsrud, and C. George. The Role of Life Events and Context in Type of Car Share Uptake: Comparing Users of Peer-to-Peer and Cooperative Programs in Oslo, Norway. Transportation Research Part D: Transport and Environment, Vol. 71, No. June 2018, 2019, pp. 186â206. https://doi.org/10.1016/j.trd.2019.01.009.

A Appendix

A.1 Levels for the one-time subscription cost and cost usage attributes in CPH

Attributes	Levels			
	RT	OWST	OWFF	P2P
One-time subscription cost	200 kr 500 kr 1000 kr	200 kr 500 kr 1000 kr	free 250 kr 500 kr	free 250kr 500 kr
Usage cost	1kr/min 4kr/min 6kr/min 200kr/6h 350kr/6h 500kr/6h 300kr/day 500kr/day 800kr/day	1kr/min 4kr/min 6kr/min 200kr/6h 350kr/6h 500kr/6h 300kr/day 500kr/day 800kr/day	1kr/min 4kr/min 6kr/min 300kr/6h 400kr/6h 550kr/6h 450kr/day 650kr/day 850kr/day	150kr/day 200kr/day 300kr/day 400kr/day 500kr/day 600kr/day 800kr/day 900kr/day 1000kr/day

Table 17: Levels for the one-time subscription cost and cost usage attributes