UNICA

UNIVERSITÀ
DEGLI STUDI
DI CAGLIARI

PH.D. DEGREE IN

Civil Engineering and Architecture

Cycle XXXV

TITLE OF THE PH.D. THESIS

Novel Neural Network Applications to Mode Choice in Transportation

Estimating Value of Travel Time and Modelling Psycho-Attitudinal Factors

Scientific Disciplinary Sector

ICAR/05

Ph.D. Student: Giovanni Tuveri

Supervisors: Prof. Italo Meloni
Francesco Piras (co-supervisor)
Eleonora Sottile (co-supervisor)

Final exam. Academic Year 2021/2022
Thesis defence: February 2023 Session

ACKNOWLEDGEMENTS

I would like to express my gratitude to my Ph.D. supervisor, Prof. Italo Meloni, for giving me the opportunity of pursuing this research endeavour, but also for his advice and for his encouragement.
I also want to sincerely thank Dr Francesco Piras and Dr Eleonora Sottile. Their priceless and constructive inputs allowed me to identify a clearer research path and to reach much more significant conclusions.
I would also like to thank Prof. Francisco Camara Pereira and all the people at DTU for their precious support during my visiting period at their University in Denmark.
Next, I would like to show gratitude to my family for supporting me in reaching this academic goal.
I gratefully acknowledge the Sardinian Regional Government for the financial support of my PhD scholarship (P.O.R. Sardegna F.S.E. - Operational Programme of the Autonomous Region of Sardinia, European Social Fund 2014-2020 - Axis III Education and training, Thematic goal 10, Investment Priority 10ii, Specific goal 10.5).

UNIONE EUROPEA
Fondo sociale europeo

REGIONE AUTONOMA DELLA SARDEGNA

FSE 2014-2020
POR
SARDEGNA

CONTENTS

1 Introduction 1

2 Literature review 5

2.1 Machine learning algorithms 5

2.1.1 Artificial Neural Networks 5

2.1.2 Bayesian Algorithms 6

2.1.3 Clustering 6

2.1.4 Decision Trees 6

2.1.5 Support Vector Machines 7

2.1.6 Ensemble Learning 7

2.1.7 Prior domain knowledge 8

2.2 Choice modelling in transportation 8

2.3 Machine learning applications for mode choice in transportation 11

2.3.1 Bayesian Algorithms, Decision Trees, and Support Vector Machines 11

2.3.2 Neural Networks 14

2.3.3 Ensemble Methods 20

2.3.4 Several Algorithms 22

2.3.5 “Hybrid” Methods 25

3 Using artificial neural networks to estimate the value of travel time in the context of mode choice modelling 28

3.1 Introduction 28

3.2 Methodology 30

3.3 Data analysis 34

3.3.1 Swissmetro 34

3.3.2 Svolta Cagliari 36

3.4 Model specification 39

3.4.1 Variable selection 39

3.4.2 Hyperparameter calibration 40

3.5 Results 41

3.5.1 Swissmetro 41

3.5.2 Svolta Cagliari 45

4 A novel application of psycho-attitudinal variables to neural network for choice modelling in transportation 49

4.1 Introduction 49

4.2 Methodology 51

4.2.1 The ICLV model 52

4.2.2 Using neural networks to reproduce the ICLV framework 55

4.2.3 Disaggregate indicators 56

4.3 Data analysis 58

4.4 Model specification 61

4.5 Results 63

4.5.1 Model 1 - LV1 (PBC) 63

4.5.2 Model 2 – LV2 (Intentions) 68

4.5.3 Model 3 - LV1 (PBC) and LV2 (Intentions) 72

4.6 Discussion 78

5 Conclusions 81

References 86

List of Tables 97

List of Figures 98

Appendix A 99

Appendix B 105

Appendix C 108

1 INTRODUCTION

Whenever researchers wish to study the behaviour of individuals challenged by choosing among a set of alternatives, they will traditionally rely on models based on the random utility theory (RUM) by McFadden, which postulates that the single individuals modify their behaviours so that they can reach the maximisation of their utility [1]. These models, often identified as discrete choice models (DCMs), usually require the definition of the utilities for each alternative, by first identifying the variables influencing the decisions of the individuals. The multinomial logit (MNL) model is the simplest and one of the most used choice models, and it is based on the independence of irrelevant alternatives (IIA) [2]. While the simplicity of the MNL models can be an advantage (*e.g.,* easier model definition, faster results estimation) it also prevents the model from reproducing complex phenomena, such as random taste variations, unrestricted substitution patterns, and correlation in unobserved factors over time. That is why researchers tried to expand the MNL modelling framework, and often resort to more advanced models, like the nested logit (NL) model, which allows to account for dependence between alternatives [3], the multinomial probit (MNP) model, which assumes that the utility functions follow a joint multivariate normal distribution [4], or the mixed logit model, whose probabilities are calculated as the integrals of MNL probabilities over a density distribution of parameters [5].

Traditionally, discrete choice models focused on directly observable variables (*i.e.*, attributes of the available alternatives and socioeconomic characteristics of the individuals) and treated users as optimizing tools with predetermined desires and needs [6]. However, such an approach is in contrast with the results from studies in social sciences which have shown that choice behaviour can also be influenced by psychological factors such as affections, attitudes, norms, and preferences [7]. Thus, recently there have been more modern formulations of DCMs which include, among their explanatory attributes, latent constructs for capturing the impact of subjective factors. These are called hybrid choice models (HCM) or integrated choice and latent variable (ICLV) models [8–10].

However, even in their most complete and overinclusive definition, DCMs are not exempt from issues, like, for example, the fact that researchers have to choose the variables to include in the model and the relations to combine them with each other to define the utilities. If this process is not implemented with the due precautions and care, to avoid human error as much as possible, it could lead to incorrect results and inaccurate interpretations of the phenomena, besides producing unreliable predictions [11].

This is probably one of the reasons which has recently lead to an influx of numerous studies using machine learning (ML) methods to study mode choice in the specific field of transportation [12], in which researchers have tried to find alternative methods to analyse travellers' choice behaviour. Machine learning models were first introduced with the current nomenclature by Samuel in the 1959, who thought of a method to program a computer to play checkers better than its programmer [13], even though some concepts had already been theorized in the previous years by McCulloch and Pitts (1943) [14], Turing [15], Minsky and Edmonds (1951) [16]. A ML algorithm is any generic method that uses the data itself to understand and build a model (it "learns" from the data), improving its performance the more it is allowed to learn [17]. This means they do not require any a priori input or hypotheses on the structure and nature of the relationships between the several variables used as its inputs.

ML algorithms can either be supervised, when the training data includes a set of labels for all the entries, that the learning system associates with the training examples [18]; or unsupervised, when the external guidance given by labelled data is missing, and building a model is more complicated [19]. The most known supervised ML algorithms are probably Artificial Neural Networks (ANNs), which are classifiers that mimic the network structure of the brain [20]. There can be several possible versions of ANN structures, but they usually include a number of hidden layers and number of nodes in each hidden layer describing the structure of the network [21]. Instead, a widely used unsupervised ML algorithm is clustering, which is defined as the issue of finding homogeneous groups (clusters) of data points in any given data set. Each cluster is then identified by a region of the multi-dimensional space defined by the data [22].

Other ML classifiers are: Bayesian estimators, a class of algorithms based on Bayes Theorem of Conditional Probability [23], that use this premise to identify to which class a feature is more likely to belong [24]; Decision Trees (DTs), classifiers which sort the input data into different groups using a set of sequential splits using a tree-like structure, where each split of the data leads to the maximum reduction in the randomness of the data at that point [25]; Support Vector Machines (SVMs), which use a kernel to transform the data into a high-dimensional space, and then find the optimal hyper-plane which divides the data into two classes [26].

ML models, in their base definition, are usually considered black-box methods, but whenever researchers felt the need for interpretability of ML results, they tried to find alternative ways to use ML methods [27]. One such method, to make ML algorithms more specific, is that of building them by using some *a priori* knowledge to induce a specific constrain on the measure to be inferred [28].

Some researchers also either transformed the outputs of machine learning algorithms so that they could be interpreted from an economic point of view [29–31], or built hybrid ML-DCM versions so that the results of their models could be interpreted like those of discrete choice models [32–35].

The object of this thesis is that of investigating the benefits and the disadvantages deriving from adopting either discrete choice models or machine learning methods to study the phenomenon of mode choice in transportation. The strongest feature of DCMs is the fact that, since they are based on behavioural theories and statistical methods, they produce very precise and descriptive results, allowing for a thorough interpretation of their outputs. This is probably an undisputable advantage they hold over ML algorithms, which often struggle to produce such informative results. On the other hand, ML models offer a substantial benefit by being truly data-driven methods and thus learning most relations from the data itself. As a matter of fact, DCMs, in the case where the base behaviour assumptions used to build the models are partially or completely incorrect, might lead to not only inaccurate but also misleading results. This should not happen with ML models, since they are built considering less restricting initial conditions.

To this end, we started our work by conducting a literature review of the state of the art of choice modelling in transportation, with a general outline of the various existing ML algorithms, and an up-to-date review of the applications of ML to choice modelling in transportation. From this literature review, a series of issues were identified, for which additional research would be benefitting to the improve this field of studies. First, there is a shortage of studies trying to extrapolate interpretable information (*e.g.*, elasticities, value of travel time, effects of latent variables) from ML models. Also, there is a lack of studies comparing this kind of results obtained by using different datasets. Finally, few studies still consider among latent variables in the specifications of ML algorithms, and even less use psycho-attitudinal indicators among their inputs.

As a first contribution, we tested an alternative method for calculating the value of travel time (VTT) through the results of machine learning algorithms. VTT is a very informative parameter to be considered, since the value that people place on saving total travel time is one of the most important indices that can be inferred [36]. As a matter of fact, the time consumed by individuals whenever they need to travel (with any mode or vehicle) normally represents an undesirable factor, thus they are usually willing to exchange their money to reduce travel times [37]. Obtaining VTT through pure ML methods is an argument which has been studied by few researchers [30, 38]. The method here proposed is independent from the mode-choice functions, so it can be applied to econometric models and ML methods equally, if they allow the estimation of individual level probabilities.

Then, another contribution of this thesis is an alternative ML method for the estimation of choice models with latent variables as an alternative to discrete choice models. This issue arose from wanting to include in ML models not only level of service variables of the alternatives, and socio-economic attributes of the individuals, but also psycho-attitudinal indicators, to better describe the influence of psychological factors on choice behaviour, not unlike it happens in ICLV models. There have also been some attempts at producing latent variable ML models [35, 39, 40], however these were limited to only using the socio-economic attributes of the individuals to build the latent factors.

The results were estimated by employing two different datasets. The first dataset (Swissmetro), which was only used to estimate the VTT, comes from a stated preference survey conducted in Switzerland in 1998. The second one, used in both implementations, was collected with a revealed preference survey conducted in 2019-2020 in Cagliari (Italy). Since neural networks results are dependent on both the values of their hyper-parameters and on their initialization, in all the applications several NNs were estimated by trying different sets of hyper-parameters to find the optimal values, and then those values were used to verify the stability of the results with different random initializations.

The thesis is organized as follows: the first part (Chapter 2) contains the literature review. In Chapter 3 we present the alternative method for calculating the value of travel time (VTT) through the results of machine learning algorithms. In Chapter 4, the method for the estimation of choice models with latent variables as an alternative to discrete choice models is described. Finally, Chapter 5 shows the conclusions reached, with the discussion of the results obtained from application of the two different ML frameworks.

2 LITERATURE REVIEW

2.1 Machine learning algorithms

This section will give an overview of the main categories of machine learning algorithms currently available. The following list just wants to give a general picture of the state of the art rather than an exhaustive one, also considering the fact that this research field is constantly evolving, and new findings are commonly found in the most up-to-date literature.

2.1.1 Artificial Neural Networks

Artificial Neural Networks (ANNs) represent a family of classifiers which mimic the network structure of the brain [20]. There are several possible versions of ANN structures for dealing with many different input data types (*e.g.* images, time-series, natural language, *etc.*), but mode choice applications usually rely on the Feed-Forward Neural Network (FFNN) (also known as the Multi-Layer Perceptron (MLP)) [12]. A FFNN consists of multiple layers of nodes (also called neurons), which include:

- an input layer, which passes the values of the features to the network;
- any number of hidden intermediate layers;
- an output layer, which returns the values predicted by the whole network.

The number of nodes in the input and output layers is fixed by the number of features and classes in the data respectively. The number of hidden layers and number of nodes in each hidden layer are hyper-parameters which describe the structure of the network [21]. Each node needs an activation function, which determines the output of that node from the weighted sum of its inputs.

There are many possible activation functions used in practice, including linear, sigmoid, tanh, softmax, rectified linear unit (ReLU), and exponential linear unit (ELU) [36]. The output values of the ANN go through a softmax function to generate classification probabilities and the weights for each link in the network are fitted to the input data. Each time the model sees all the data once is defined as an epoch. The number of epochs can be limited to regularise the model and limit overfitting issues. In a fully connected network, every node in one layer is linked to every node in the next layer. Further regularisation can be applied using the “dropout” hyper-parameter, which specifies a portion of the neurons to be dropped randomly from the network for each batch of data [37].

2.1.2 Bayesian Algorithms

The Bayesian estimators are a class of algorithms based on Bayes Theorem of Conditional Probability [23], and use this premise to identify to which class a feature is more likely to belong. Among them, Naïve Bayes classifiers assume a great level of simplification by considering that all features are independent from each other, but at the same time they are show remarkably good results in practice, often comparable to those of much more sophisticated techniques [24]. Bayesian Network classifiers, instead, are representations of probability distributions that generalize the naive Bayesian classifier. They are built as directed graphs that allow to represent the joint probability distribution over a set of random variables. Each vertex in the graph corresponds to a random variable, while the edges represent direct correlations between those variables [38].

2.1.3 Clustering

The clustering problem is basically defined as the issue of finding homogeneous groups of data points in any given data set. Each of these groups is then called a cluster, which be defined as a region of the multi-dimensional space defined by the data in which the density of objects is locally higher than in other regions. The simplest form of clustering is partitional clustering which aims at partitioning a given data set into disjoint clusters so that specific clustering criteria are optimized [22]. Among the many clustering algorithms, the most used are perhaps *k-means clustering*, a simple procedure which allows to obtain a set of partitions which are reasonably efficient in the sense of within-class variance [39]; and hierarchical clustering [40], which starts with a separate cluster for each of the distinct points of the dataset, for then agglomerating the two closest clusters (according to a chosen metric) sequentially until all the points belong to one hierarchically constructed cluster [41].

2.1.4 Decision Trees

Decision Trees (DTs) (also known as Classification and Regression Trees - CARTs) are classifiers which sort the input data into different groups using a set of sequential splits using a tree-like structure. Most commonly, Decision Trees are fitted using recursive binary splits, where each split of the data is chosen so that it leads to the maximum possible reduction in the randomness of the data at that point. The metrics which are generally used to measure how much the data are shuffled, are Gini impurity and entropy. To calculate a split, each possible binary split (the value for the data is less/greater than a certain threshold value) is tested for each feature. The split point which leads to the greatest reduction in the impurity or entropy (across all features) of the data is then selected, resulting in two new child nodes. The same procedure is then applied recursively to each child node, until a stopping condition (set when choosing the hyper-parameters of the algorithm) is met.

For example, setting a value for the maximum depth of the tree specifies the maximum number of sequential splits which can be encountered along a path from the first node; setting the minimum leaf size specifies the minimum number of data points that each child node must include after a split in order for the same split to happen; setting a minimum split size specifies instead the minimum number of samples a node must include before a split for that split to be considered possible at that node [25].

2.1.5 Support Vector Machines

The Support Vector Machine (SVM) algorithm uses a kernel to transform the data into a high-dimensional space. The algorithm then finds the optimal decision surface (or hyper-plane) in the transformed space which divides the data into two classes. For linearly separable data (within the transformed space), the optimal hyperplane exactly divides the data without any misclassification while also maximising the possible margin. The margin is, by definition, the perpendicular distance between the hyperplane and the nearest data points (these data points are called support vectors). Instead, for complex, real-world examples, the input data are usually not linearly separable, even within the transformed space. There needs to exist a balance between the width of the hyperplane and the number of misclassifications of the training data and this is controlled by using a regularisation parameter (C). A higher value of C corresponds to a higher importance given to the misclassified points (higher variance), while a lower value of C will give more importance to the width of the hyperplane (higher bias). While Support Vector Machines are inherently binary classifiers, they can also be used for multiclass classification using either a one-vs-rest or one-vs-one strategy. SVMs output a continuous score for each prediction, which can be interpreted as the confidence of the classification [26]. There are several specifications of the kernels which can be used to transform the data, including linear (*i.e.* no transformation), polynomial or Radial Basis Function (RBF) (Gaussian or exponential) [42].

2.1.6 Ensemble Learning

Ensemble Learning (EL) (or Ensemble Methods) combines several individual predictive models (called estimators) in an ensemble to improve the quality of predictions compared to the results of single estimators. This result is based on the fact that, provided the errors of the estimators can be considered independent (*i.e.* the learners are uncorrelated), and the individual models are more likely to be right than wrong, then combining them in an ensemble reduces their individual uncertainty [43].

Many meta-algorithms that can combine estimators exist. There are some algorithms which train estimators on the data in parallel, *e.g.* Bootstrap Aggregating (Bagging) [44] and Random Forest (RF)

[45], as well as algorithms where the single weak learners are estimated sequentially, *e.g.* AdaBoost [46] and Gradient Boosting (GB) [47].

DTs are the most used estimators for Ensemble Learning, but they also usually present high variances, making them unstable (a small variation in the input results in large differences in the output). Although this also means that it is relatively easy to train uncorrelated DTs compared to more stable classifiers. Also, DTs are algorithmically simple to fit and obtain predictions from, meaning that large ensembles of DTs can be fit and predicted in a reasonable amount of time. For ensembles of discrete classifiers, probability-like values can be obtained by calculating the proportions of each class prediction across the estimators in the ensemble. For Gradient Boosting Decision Trees (GBDT), the DTs in the ensemble are trained to output discrete regression values. These values are then passed through a softmax function to output choice probabilities. One of the main hyper-parameters of EL algorithms is the number of estimators in the ensemble. In parallel approaches, this number must be specified, while for sequential approaches a stopping criterion can be applied based on out-of-sample predictive performance (not unlike the number of epochs in ANNs) [12].

2.1.7 Prior domain knowledge

Machine learning algorithms can usually be seen as general-purpose tools, which "learn" from the data without using any assumptions about the possible relations existing between the different variables, doing so also independently from the very nature of the data itself. A more specific and specialized class of algorithms can be built by relying upon *a priori* domain knowledge to constrain the concept to be inferred [28]. As a matter of fact, the training dataset samples are usually unable to fully describe the dynamics to be simulated in a way to enable the use of ML models alone. Supporting a ML model with prior domain knowledge on the specific field to which the data belongs can be thought of as filling the gaps in the same observed data with the knowledge of the internal functioning of the system. This approach often leads to a reduction of the size of the search space during training, simplifying the whole process [48]. The models thus become "physically consistent", in the sense that they gain a sufficient grounding in physical principles (reached by the use of physical laws or empirical algorithms) [49].

2.2 Choice modelling in transportation

The most common theoretical construct for building discrete choice models in transportation is the random utility theory (RUM) by McFadden, which postulates that the decision makers (single individuals) base their behaviour on the maximisation of their utility. The modeler cannot however fully observe the decision maker's utility, but only some attributes (*e.g.*, transport alternative characteristics, personal preferences, sociodemographic levels) can be known. Therefore, the modeler has to assume that the utility is composed by an observed part (V_{qi}, function of the known attributes) and a random residual part (ε_{qi}), which explains the influence of all unobserved factors [1].

The multinomial logit (MNL) model is the simplest and one of the most used choice models. The logit model is obtained by assuming that each ε_{qi} is an independently, identically distributed extreme value, and that the choice probability for alternative i and individual q is given by the relation $P_{qi} = e^{V_{qi}} / \sum_j e^{V_{qj}}$. The multinomial logit model is based on the independence of irrelevant alternatives (IIA), meaning that the relative probability of choosing one alternative over another is independent from the other alternatives [2, 50].

To overcome the IIA limitation, a class of models known as generalized extreme value (GEV) models was developed by McFadden (1978). The nested logit (NL) model is one of the more commonly used models in this class. The idea behind a nested logit is to divide the choice set into nests of alternatives, in such a way that for any two alternatives that are in the same nest, the ratio of probabilities is independent of the attributes or existence of all other alternatives and for any two alternatives in different nests, the ratio of probabilities can depend on the attributes of other alternatives in the two nests [3, 51].

An alternative solution to overcome the limits of IIA is to use the multinomial probit (MNP) framework. The MNP model assumes that the utility functions follow a joint multivariate normal distribution with zero mean and arbitrary covariance matrix. This means that the variances may be different, and the error terms may be correlated in any fashion, and probit models can allow any pattern of substitution and handle panel data. The only limitation of probit models is that the normal distribution assumption for model parameters might be inappropriate in some situations and can lead to issues in results interpretation. This can however be solved by using a Generalized Multinomial Probit model with truncated normal random parameters [4, 52, 53].

The issues of IIA can also be resolved by using a heteroscedastic extreme value model, which allows a more flexible structure among alternatives than nested logit model, while also requiring a much lower computational burden than the multinomial probit model. The heteroscedastic extreme value model allows the amount of stochasticity of the utility of alternatives to differ. That is, it allows different variances on the random components across alternatives, which is likely to occur when the variance of an unobserved variable that affects choice is different for different alternatives [54].

The mixed logit model is a generalization of the standard logit model and can approximate any random utility model to various degree of accuracy. Mixed logit choice probabilities are calculated as the integrals of multinomial logit probabilities over a density distribution of parameters that can vary randomly across individuals [5].

Many transportation applications involve the assessment of influences on a choice amongst ordered discrete alternatives, like situations where respondents are asked to provide ratings, ordered opinions or in general any categorical preference. Although these response data are discrete, use of standard or nested multinomial discrete models is not a proper way to model data of an ordered nature. To address the problem of ordered discrete data, ordered probability models have been developed. An ordered-response model postulates the presence of a latent continuous variable for each individual q. [55, 56]

A limitation of the standard ordered probit / logit is that it is based on the assumption that the thresholds between categories are the same for every individual in the sample. This can lead to biased and inconsistent estimates of the effect of variables. Instead, models in the class of generalized ordered probit/logit (also known as hierarchical ordered probit/logit) allows different thresholds to be specified based on the variables, meaning they can also be different between categories [56, 57].

A multivariate ordered-response framework can be used to account for possible within-outcome correlations. A multivariate ordered response model assumes the presence of a set of multivariate continuous latent variables whose partitioning maps into the observed set of ordered outcomes. Such a system allows to use a general covariance matrix for the underlying latent variables, which can also mean that there exists a flexible correlation between the observed ordered outcomes [56–58].

Microeconomic theory tends to consider decision makers as rational actors constantly evaluating the costs and benefits associated with any choice, with the aim of maximizing their personal utility. Traditional discrete choice models focus their analysis on observable variables, such as the attributes of the alternative and socioeconomic characteristics of the individual, and treat users / consumers as optimizing tools with predetermined desires and needs [6]. However, such an approach clashes with the findings from studies in social sciences which have shown that choice behaviour can also be influenced by psychological factors such as affections, attitudes, norms, and preferences [7]. Only recently models of disaggregate decision-making which include latent constructs for capturing the impact of subjective factors on choice process have started to be considered. These models are called hybrid choice models (HCM) or integrated choice and latent variable (ICLV) models [8, 9].

In the general formulation of the ICLV model, there are two distinct components: a latent variable model and a discrete choice model. The latent variable model is in turn composed of a set of structural equations, describing the latent variables in terms of observable variables, and a set of measurement equations, linking the latent variable to indicators. Since latent constructs cannot be observed, the analyst obtains information about them from indicators, *i.e.*, observed responses to questions of a survey. By integrating discrete choice and latent variable models, the latent variables can be seen as explanatory variables included in the utilities of choice alternatives [10].

The recent increase in quantity and quality of measurement tools and technologies capable of recording enormous amounts of data (big data), has allowed the production of datasets often characterized by complex interdependent structures (multivariate mixed data). Bhat proposed a different way of treating mixed data by using a Generalized Heterogeneous Data Model (GHDM). This model is an evolution of the Integrated Choice and Latent Variable models able to handle mixed types of dependent variables (nominal, ordered, continuous, count) by representing the covariance relationships among them through a reduced number of latent factors. GHDM models consist of two components: a latent variable structural equation model (SEM), and a latent variable measurement equation model (MEM). In the SEM, latent variables relevant to the outcomes of the MEM are hypothesized, based on theoretical psychological considerations and earlier qualitative/quantitative studies. In the MEM, endogenous variables are described as functions of latent variables and exogenous variables [59].

2.3 Machine learning applications for mode choice in transportation

Several examples of successful applications of machine learning algorithms for modelling mode choice in transportation are available in the literature. To improve clarity and readability, the following analysis is split by considering the main techniques studied by the researchers in their papers. Additional information on these examples, including the year of the studies, sample sizes, data sources, and territorial contexts where the data was collected, is reported in additional tables placed after each paragraph (from Table 1 to Table 6).

2.3.1 Bayesian Algorithms, Decision Trees, and Support Vector Machines

Van Middlekoop *et al.* [60] tried to show that tourists do not necessarily maximize their utility in selecting a travel mode and their choice behaviour is context dependent. They used a DT-like algorithm (specifically CHAID - chi-squared automatic interaction detection), and their results indicated how the proposed methodology can be applied successfully to better understand choice behaviour.

Zhang and Xie [61] applied for the first time a SVM to travel mode choice modelling and compared it with a MNL model and an ANN. The SVM performed better than the other two models in predicting choices. The overall fitting performance of the ANN was the best, but at the same time its prediction performance was the worst. All three models could produce good predictions for modes with few observations.

Zenina and Borisov [62] investigated the performance of mode choice analysis with different classification methods - DT, discriminant analysis and MNL. They initially found 67% - 78% of correctly classified instances using DTs, but, after additional data pre-processing, they managed to improve this result reaching a percentage of correctly classified instances of 80% - 92%, which was better than the other methods analysed.

Ma [63] tried capturing the non-linear causal effects of related determinants on individuals' mode choice behaviour by using a rule-based approach based on Bayesian Networks. The results were compared with a MNL, and they found that the performance of the two approaches showed similar corrected prediction rates.

Pitombo *et al.* [64] applied a two-step method to estimate mode choice, first using a DT to select the attributes which most influenced mode choice, then comparing the performance of the DT with a MNL and applying an Ordinary Kriging to predict the mode choice. The DT showed its effectiveness since its comparison with the MNL indicated that the variables chosen by both the methods are the same, with the advantages of the straightforward application of DTs.

Tang *et al.* [65] used a series of DTs to explore the underlying rules of travellers' switching decisions between two modes under a proposed framework of dynamic mode searching and switching and compared the results with (unspecified) logit models. The DTs they built were able to correctly classify the minority mode with a high overall prediction accuracy. In many cases DTs outperformed logit models in both individual prediction level and aggregate prediction level.

Semanjski *et al.* [66] developed an SVM-based model to predict future mobility behaviour from crowdsourced data collected through a dedicated smartphone app. The SVM model had a success rate of 82% in forecasting the mode of transportation to be used for the next trip.

Brathwaite *et al.* [67] began bridging the gap between machine learning methods and economic theories of human decision making by providing a microeconomic framework for interpreting a Bayesian DT and testing this method on an application of bicycle mode choice. The proposed DT was more than 1,000 times more likely to be closer to the application true data-generating process than a MNL model and provided forecasts consistent with observed mode share.

Ma *et al.* [68] modelled commuters' travel mode choice by using a two-stage procedure, first drawing relevant prior partial relationships between variables, and using them as structure constraints in a structure learning task of Bayesian Networks, then using a model averaging approach to obtain a statistically sound Bayesian Network. The methodology proved to be a valuable supplemental analytical tool to conventional travel choice forecast models to identify impacts of providing more information about key variables.

Ding *et al.* [69] applied a gradient boosting DT to examine the influences on commute mode choice of commuting programs and built-environment characteristics at both residential and workplace locations. The model was able to predict a large influence of built environment on car mode choice. The lack of a benchmark or comparison with other models does not allow to comment on the model overall performance.

Zhu *et al.* [70] proposed a new modelling approach that utilizes a mixed Bayesian Network (BN) for travel decision inference. One mono-dimensional BN, two bi-dimensional BNs, one DT and two nested logit models were developed for comparison. The results indicated that bi-dimensional BNs had a more stable performance in the testing dataset than the mono-dimensional BN. The bidimensional BN approach gave better accuracy than the DT and the NL models, especially for small probability alternatives.

Jing *et al.* [71] analysed the travel selection behaviour of Chinese high-speed train passengers by using Support Vector Machine, Nested Logit models, and Multinomial Logit models. The SVM resulted the most accurate, followed by the NL.

Pirra and Diana [72] proposed a new approach based on SVM for recognizing travel mode choice patterns. Although they discovered that the resulting accuracy was not comparable to that of a "good" model, they also realized SVM could give a first approximation in case studies were large amounts of data need to be processed quickly and heuristic solutions are acceptable.

Table 1. Details for the literature regarding Bayesian Algorithms, Decision Trees, and Support Vector Machines

Authors	Year	Data source	Territorial context
Van Middelkoop, Borgers, and Timmermans [60]	2003	**3,562 observations** Vacation behaviour, sociodemographic, and vacation-related variables of a panel representative of the Dutch population.	Netherlands
Zhang and Xie [61]	2008	**5,029 observations** Home-to-work commute trip records, sociodemographic information, attributes of each travel mode, land use, and other related data.	San Francisco (USA)
Zenina and Borisov [62]	2011	**7,171 (500) observations** Personal interviews which collected information on socioeconomic and demographic variables, travel characteristics, and travel influence conditions.	*Not available*
Ma [63]	2015	**7,235 observations** Mobility survey of cross border workers in the Greater Region of Luxembourg.	Luxembourg
Pitombo, Salgueiro, Da Costa, and Isler [64]	2015	**2,791 observations** Household interview from an Origin–Destination and data from the water supply database used to obtain the geographical coordinates.	São Carlos (Brazil)
Tang, Xiong, and Zhang [65]	2015	**72,536 observations** Data set that has been geocoded and includes household data, person data, vehicle data and daily trip data.	Washington DC (USA)
Semanjski, Lopez, and Gautama [66]	2016	**17,040 observations** Data on mobility behaviour collected using an Android smartphone application over four months.	Flemish-Brabant province (Belgium)
Brathwaite, Vij, and Walker [67]	2017	**1,015 observations** Household Travel Survey from California with observations representing home to work or school commute tours.	Oakland, Berkeley, San Francisco (USA)
Ma, Chow and Xu [68]	2017	**5,040 observations** Mobility survey of the cross-border workers in Luxembourg, containing one-day travel diaries with related spatial and socio-demographic information, among other variables.	Luxembourg
Ding, Cao, and Wang [69]	2018	**6,392 observations** Regional household travel survey which includes socio-economic and demographic characteristics, and work-related attributes.	Washington DC (USA)
Zhu, Chen, Xiong, and Zhang [70]	2018	**5,213 observations** Travel diaries that documented the activities of all household members on an assigned day and detailed information including household, person, vehicle and trip information.	Washington, Baltimore (USA)
Jing Liu, Zhang, and Su [71]	2019	**857 observations** Questionnaire distributed every day for 1 week, from 8:00 a.m. to 8:00 p.m., to passengers travelling between two places from Beijing South Railway Station.	Beijing (China)
Pirra and Diana [72]	2019	**39,167 observations** National Household Travel Survey which provides information on daily trips in the United States, whose data was improved with interviews gathered in New York State.	New York State (USA)

2.3.2 Neural Networks

Perez and Pietrzyk [73] were among the first researchers to suggest that ANNs should be tested for mode choice.

Raju *et al.* [74] studied the applicability of an ANN for modelling modal choice, using a relatively small dataset (535 observations) from a household survey in Guwahati (India). They managed to demonstrate how the ANN was capable of accurately (85%) capturing the patterns in the learning process, even though these results were obtained by simplifying the data and considering only two modes. However, this study opened the path for many future developments and improvements.

Subba *et al.* [75] explored the use of an ANN for mode choice modelling and compared it to a MNL, finding the performance of the former to be much superior both in calibration and in prediction.

Hensher and Ton [76] also analysed the merits of ANNs by comparing their predictive capability and those of nested logit models in the context of commuter mode choice. However, they found no clear indication of which one of the two approaches was better.

Cantarella and de Luca [77] tried showing that ANNs can be an effective tool for travel demand analysis, introducing a new architecture which included one extra layer for perceived utility. Their results showed that the proposed ANN model was feasible, and the most performing ANN could be considered as a good approximation of the best devisable choice model.

Cantarella and de Luca [78] also described the application of ANNs (specifically a multilayer feedforward ANN) to support travel demand analysis, showing that they can be applied to analyse transportation mode choice. They then compared them with various RUMs (dogit, MNL, CNL), and the obtained results showed that ANNs turned out to be rather effective and they may outperform RUMs when the values of mode shares are quite similar.

Chalumuri *at al.* [79] validated and compared mode choice models based on Logit and ANNs for work, education and other-purposes trips. From the comparison of the results, it can be concluded that the models were comparable and considered to be consistent in predicting the choice behaviour. However, it should be noted that they only considered a simple Logit model, which might result in an over-simplification of the representation, since more advanced and reliable models were available.

Zhao *et al.* [80] applied a probabilistic ANN for travel mode choice modelling. They found the ANN to be valid in structure simplification and excellent in training time reduction, and its prediction results have proved to be more accurate than those of DCMs.

Yin and Guan [81] established a traffic mode choice model based on a back-propagation ANN and verified the model by using data from a travel survey of residents in Jinan City (China). The result showed that the model had a very good practicality and could be utilized for predicting resident trips mode choice.

Gao *et al.* [82] used a MNL model to demonstrate that transit network layout has significant effect on resident mode choice. Based on parameter estimation, the factors affecting mode choice were further screened and then regarded as the input data to a back propagation ANN for training and forecasting, to further confirm their findings.

Omrani *et al.* [83] used an evidential ANN (ENN) to support management decision making and to build predictions under uncertainty related to changes in people's behaviour, in the economic context, or in the environment and policy. The rates of successful prediction obtained by the ENN and several alternative ANN approaches were compared by cross-validation. The results showed that the performance of the ENN supported its use as an alternative procedure for modelling travel mode choice.

Omrani [84] used two different ANN specifications (multiple layer perceptron and radial basis function), MNL and SVMs for predicting travel mode of individuals. They then compared the rates of success for predicting the travel mode choice using cross-validation techniques. They found the average probability of correct assessment of SVM and MNL was higher, however the ANNs had slightly better performance.

Hussain *et al.* [85] compared two mode choice models, MNL and ANN, with the purpose of evaluating the accuracy levels in the predictability in each model. In both terms of predictability and validation, the ANN exceeded the MNL results.

Assi *et al.* [86] made a comparison between the efficiency and robustness of the Logit regression model and a multilayer perceptron ANN to predict and explain the mode choice behaviour of high school students. The results guaranteed that ANNs will perform better when it comes to understanding the predictive power of their mode choice behaviour.

Golshani *et al.* [87] compared the performance of MNLs with the performance of ANNs in the contexts of trip departure time and mode choice behaviour. The ANN model outperformed the statistical models in terms of implementation burden and prediction accuracy (87% and 64% of correct predictions, respectively).

Lee *et al.* [88] investigated the capabilities of four types of ANN model (backpropagation, radial basis function, probabilistic, clustered probabilistic) and compared their prediction performance with a conventional MNL for mode choice problems. The cross-validation results revealed that the four ANNs achieve better prediction accuracies (around 80%) than the MNL (around 70%), with the clustered probabilistic ANN showing the highest performance.

Minal and Sekhar [89] modelled the mode choice of commuters using two discrete choice model (MNL and nested logit) and a non-conventional machine learning method (ANN). A comparative evaluation of the results showed that the model developed by using an ANN is the superior of the three due to higher accuracy and better exploratory power.

Aschwanden *et al.* [90] presented a novel ANN-based modelling technique capable of predicting transportation mode distribution from georeferenced trips and satellite images. The ANN was able to identify the urban patterns that are more conducive for different modes of transportation, despite the limited information density provided by satellite images.

Ma and Zhang [91] proposed a deep ANNs with entity embeddings to jointly learn meaningful representations of categorical variables and accurate travel mode predictions. The entity embedding technique turned out to be good for enhancing the prediction performance and it could boost the performances of DT-based models.

Wang *et al.* [29] demonstrated the use of behavioural knowledge for the design of a deep ANN architecture with alternative-specific utility functions (ASU-DNN), improving both the predictive power and interpretability. The results demonstrated how behavioural knowledge can function as an effective domain-knowledge-based regularization method, and how it could improve the power of ANNs in choice analysis.

Wang *et al.* [30] provided an empirical method of numerically extracting from ANN results valuable economic information such as choice probability, probability derivatives (or elasticities), and marginal rates of substitution such as value of travel time. ANN-based choice models generated reasonable economic information at the aggregate level. They resulted in roughly S-shaped choice probability curves and inverse bell-shaped probability derivatives.

Zhang *et al.* [92] proposed a deep ANN framework for traffic mode choice in which a local-connected layer extracted a utility specification from the data, and then, a fully connected layer augmented feature representation. The first local-connected hidden layer partially replaced the manual utility specification and the second fully connected hidden layer enabled the model to eliminate the IIA problem.

Buijs *et al.* [93] investigated an approach where the travellers' transportation mode was predicted through an ANN trained on choice sets and user specific attributes inferred from data of the Amsterdam metropolitan area. The models showed better results when predicting the choice of mode for trips taking place on the same network as the training data.

Li *et al.* [94] conducted a comparative analysis of regression-based multinomial models and ANN models in intercity travel mode choices. MNL and Bayesian multinomial logit (BMNL) were compared with the radial basis function (RBF) and multilayer perceptron (MLP). The MLP performed best in terms of predictive accuracy.

Van Cranenburgh and Kouwenhoven [95] proposed a novel ANN-based method to derive the VTT distribution without making assumptions about the shape of the distribution or the error terms. The method was both tested on a series of data derived from Monte Carlo experiments and applied to data from the 2009 Norwegian VTT study. The results were then validated by comparing them to those of discrete choice models and nonparametric methods, showing very promising results.

Table 2. Details for the literature regarding Neural Networks (part 1 of 2)

Authors	**Year**	**Data source**	**Territorial context**
Perez and Pietrzyk [73]	1995	**~27,000 observations** Data were collected from several thousand companies, some of which had implemented strategies to increase their average vehicle ridership.	Los Angeles (USA)
Raju, Sikdar, and Dhingra [74]	1996	**535 observations** Socioeconomic and travel-related information about the trip-makers from household survey data, transport network data.	Guwahati (India)
Subba Rao, Sikdar, Krishna Rao, and Dhingra [75]	1998	**4,335 observations** Details of access legs of work trips, socio-economic characteristics of the traveller, system characteristics of the alternatives.	Mumbai (India)
Hensher and Ton [76]	2000	**801 observations** Data from a stated choice experiment examining the potential impacts of transport policy instruments on reductions in greenhouse gas emissions.	Sydney, Melbourne (Australia)
Cantarella and de Luca [77]	2003	**2,808 observations** Journeys of students towards the country-side location of the University of Salerno from outside the city of Salerno, obtained by interviews at parking locations.	Salerno (Italy)
Cantarella and De Luca [78]	2005	**1,067 + 2,350 observations** Journeys in the central area of the Veneto Region, and extra-urban journeys of students towards the University of Salerno, which LoS attributes were computed through transportation network models.	Veneto, Salerno (Italy)
Chalumuri, Errampalli, Bujangan, and Subamay [79]	2009	**1045 + 1018 observations** Home-interview surveys collecting the choice behaviour of the commuters for the journeys to different purposes namely work, education and other purposes.	Visakhapatnam, Nagpur (India)
Zhao, Shao, Li, Dong, and Liu [80]	2010	**967 observations** Resident trip survey along a certain suburban railway line in Beijing, covering respondents' travel information, which describes the characteristics of both traffic mode and traveller.	Beijing (China)
Yin and Guan [81]	2011	**1,007 observations** Residents travel surveys.	Jinan City (China)
Gao, Zhao, Zhuge, Zhang, and McCormack [82]	2013	**650 observations** A household survey including mode choice, individual characteristics and travel features, city features, transportation policies.	Baoding (China)

Table 3. Details for the literature regarding Neural Networks (part 2 of 2)

Omrani, Charif, Gerber, Awasthi, and Trigano [83]	2013	**9,500 observations** Data from a Socioeconomic Panel Survey containing individuals' characteristics, transportation mode specifications, and data related to places of work and residence.	Luxembourg
Omrani [84]	2015	**3670 observations** Data from a socio-economic panel from a sample of households from the resident population in Luxembourg.	Luxembourg
Hussain, Mohammed, Salman, and Borhan [85]	2017	**620 observations** A survey carried out in Baghdad in four areas with higher percentage of the population and noticeable employment of private cars.	Baghdad (Iraq)
Assi, Nahiduzzaman, Ratrout, and Aldosary [86]	2018	**597 observations** Questionnaire designed to collect information from parents about the mode choice behaviour of their children, including trip characteristics and household details.	Khobar (Saudi Arabia)
Golshani, Shabanpour, Mahmoudifard, Derrible, and Mohammadian [87]	2018	**9,450 observations** Travel Tracker Survey containing trip information as well as household and individual level sociodemographic characteristics and activity-related variables.	Chicago (USA)
Lee, Derrible, and Pereira [88]	2018	**4,746 observations** Travel Tracker Survey containing trip information as well as household and individual level sociodemographic characteristics and activity-related variables.	Chicago (USA)
Minal and Sekhar [89]	2018	**94 observations** An online survey conducted in the month of February 2013 towards the population of Delhi.	Delhi (India)
Aschwanden, Wijnands, Thompson, Nice, Zhao, and Stevenson [90]	2019	**63,365 observations** Two data sets: georeferenced trips from the Victorian Integrated Survey of Travel and Activity and satellite images from Google Maps.	State of Victoria (Australia)
Ma and Zhang [91]	2020	**81,086 observations** Traffic mode survey dataset which combines individual records of the London Travel Demand Survey with corresponding travel trajectory for all travel choices.	London (UK)
Wang Mo, and Zhao [29]	2020	**8, 418 + 2, 929 observations** An online survey data collected in Singapore with the aid of a professional survey company and a public dataset containing data about a stated preference survey in Netherlands.	Singapore; Netherlands
Wang, Wang, and Zhao [30]	2020	**8,418 observations** Stated preference survey conducted in Singapore, with questions about home and working locations, current travel mode, and seven choice scenarios varying by availability and attributes.	Singapore
Zhang, Ji, Wang, and Yang [92]	2020	**1,000,000 observations** Data from the public transportation system, taxi orders, and anonymous navigation users of Beijing, comprising the travels from a bus, subway, taxi, and private car.	Beijing (China)
Buijs, Koch, and Dugundji [93]	2021	**106,647 observations** GPS data is collected during three time periods spanning about one month each and generated alternatives for each user.	Amsterdam (Netherlands)
Li, Wang, Wu, Chen, Zhou [94]	2021	**985 observations** Questionnaires collected in Xi'an in addition to the travel distances calculated by Baidu Maps using the real route between the cities of origin and destination.	Xi'an (China)
Van Cranenburgh and Kouwenhoven [95]	2021	**52,488 observations** Norwegian 2009 VTT data set, in which the actual observed choices, were replaced with synthetically generated choices.	Norway

2.3.3 Ensemble Methods

Biagioni *et al.* [96] developed a novel adaptation of data-mining methods through the use of ensemble classifiers (based either on DTs or Naïve Bayes classifiers). By defining the notion of an “anchor mode” as the mode selected on the first trip of a tour, this ensemble was trained with and without knowledge of the anchor mode. The results of the ensemble were then compared to other machine learning classifiers and MNL. The ensemble achieved a high level of accuracy, precision, and recall, outperforming MNL.

Rasouli and Timmermans [97] explored the idea of replacing a single representation with an ensemble of models using DTs and investigated whether the use of a model ensemble would reduce errors and uncertainty in predicting transport mode decisions. They found that predictive success tends to increase when increasing ensemble size (20 or more DTs), and that the feature importance also varied based on ensemble size.

Ermagun *et al.* [98] jointly modelled decisions on escorting children and modes of transport by using a RF, and a nested logit model. They found the RF significantly outperformed the nested logit, with a 62% prediction accuracy for the former compared with 38% for the latter, Also, the RF model performed better in all categories but private cars, for which the accuracy of the two models was similar.

Sekhar *et al.* [99] modelled the mode choice behaviour of commuters in Delhi by considering a DT-based RF, and then compared it to a MNL. The RF model was superior, with higher prediction accuracy (98%) than the MNL models (77%).

Brondeel *et al.* [100] developed a simulation to evaluate the impact on physical activity of the transport mode shifts anticipated in a Urban Mobility Plan, using a simulation method based on RF models.

Wang and Ross [101] explored the application of an extreme gradient boosting (XGB) model based on DTs to travel mode choice modelling and compared the results with a MNL model. The performance of the XGB model exceeded that of the MNL not only in predicting the choices of all modes together, but also for every individual mode.

Chang *et al.* [102] employed a data fusion model based on a stacking strategy and proposed a hybrid model of the unsupervised Denoising Autoencoder (DAE) combined with a supervised RF to improve the prediction accuracy of mode choice. Compared with traditional MNLs, RF improved the classification accuracy by 27%. The DAE combined with the RF could better model the travel mode choice behaviour using a three-stage architecture.

Chapleau *et al.* [103] employed a RF to characterize the use of eight different travel modes observed in two consecutive household travel surveys. The RF produced a higher correct prediction rate and a lower total error rate than MNL, despite the simplicity of the input data structure. In addition, the application of the RF to a larger and independently collected dataset demonstrated its robustness.

Cheng *et al.* [104] proposed a robust RF method to analyse travel mode choices and examined the prediction capability and model interpretability. A comparison between the RF, SVM, AdaBoost and MNL showed how RF and SVM are the best models in prediction accuracy. However, RF is more computationally efficient than SVM because it takes less time to train the model.

Lee *et al.* [105] applied a gradient boosting machine (GBM) to develop a choice model on three choice alternatives related to autonomous vehicles, based on stated preference survey data, which also included attitudinal statements from respondents. The prediction performance of GBM was evaluated by conducting a 5-fold cross-validation and showed around 80% accuracy.

Li *et al.* [106] set up a model with a RF for the travel mode choice of passengers after an urban rail transit system was put into use, to find out the impact of different travel features. The RF algorithm could analyse the importance of passenger travel features and reached a prediction accuracy of 95.9% when travel cost was taken into consideration.

Ceccato *et al.* [107] evaluated substitution rates of car-sharing against private cars and public transport using a RF classifier and a Binomial Logit model. Binomial Logit and RF had similar predictive powers, however only RF provided a deep understanding of the effect of explanatory variables.

Kim [108] proposed an interpretable ML approach to improve their interpretability concerning travel mode choice modelling. XGB was applied and it outperformed other ML models when considering variable importance, variable interaction, and accumulated local effects.

Table 4. Details for the literature regarding Ensemble Methods

Authors	Year	Data source	Territorial context
Biagioni, Szczurek, Nelson, and Mohammadian [96]	2008	**116,666 observations** Activity-based dataset consisting of one-day and two-day activity diaries, complete with socio-demographic and trip-based attributes.	Chicago (USA)
Rasouli and Timmermans [97]	2014	**1,446 observations** Trip-diary with details about activities, in addition to individual and household sociodemographic characteristics.	Netherlands
Ermagun, Rashidi, and Lari [98]	2015	**4,700 observations** Characteristics of the built environmental, the transportation system, students' trips, and socioeconomic of households.	Tehran (Iran)
Sekhar, Minal, and Madhu [99]	2016	**5,000 observations** Travel behaviour data has been collected through predesigned questionnaire.	Delhi (India)
Brondeel, Kestens, and Chaix [100]	2017	**82,084 observations** Transport behaviour and accelerometer data collected for a study in Ile-de-France.	Ile-de-France (France)
Wang and Ross [101]	2018	**51,910 observations** Household travel survey data collected in the Delaware Valley region, including trip records, travel mode and income levels.	Delaware (USA)
Chang, Wu, Liu, Yan, Sun, and Qu [102]	2019	**52,265 observations** Travel diary survey data containing trip specific data, socio-economic data about the participants and information on households.	Germany
Chapleau, Gaudette, and Spurr [103]	2019	**72,180 + 86,836 observations** Household travel surveys which measured household attributes, individual characteristics and trip attributes.	Montreal (Canada)
Cheng, Chen, De Vos, Lai, and Witlox [104]	2019	**7,276 observations** Household surveys conducted to get resident's travel information and their socio-demographics, combined with information on the built environment.	Nanjing (China)
Lee, Mulrow, Haboucha, Derrible, and Shiftan [105]	2019	**4260 observations** Data from a survey given only to individuals who currently drive a car for their daily commute to work or school, designed to investigate individuals' likelihood of choosing their future vehicle.	Israel; North America
Li, Gao, Zhang, and Liao [106]	2020	**733,734 observations** IC card data of the bus and rail transit which includes IC card identification, transaction date, boarding and alighting time and stops/stations, and latitude and longitude information.	Xiamen (China)
Ceccato, Chicco, and Diana [107]	2021	**1,050 + 200 observations** Stated-preference travel survey and revealed-preference survey.	Turin (Italy)
Kim [108]	2021	**172,889 observations** NHTS dataset from Seoul, which includes individual travel diaries that recorded every daily trip taken.	Seoul (South Korea)

2.3.4 Several Algorithms

Xie *at al.* [109] investigated the capabilities and performance of DTs and ANNs for work travel mode choice modelling, comparing them to a traditional MNL model. The comparative evaluation showed that the two machine learning models had slightly better prediction capabilities than the MNL, and that the NN outperformed the other two models.

Shukla *et al.* [110] proposed a novel data-driven methodology to address some issues identified in DCMs using ANNs and DTs combined with fuzzy datasets. The results from the various analysis they conducted suggested that the use of fuzzy sets and a tour-based model for mode choice achieved high performances. However, they decreased the value of their findings by not including a comparison with DCM, since they explicitly expressed their aim was to address some of the issues of such models.

Hagenauer and Helbich [111] compared the predictive performance of MNL and several machine learning classifiers (Naïve Bayes, SVM, ANN, Boosting, Bagging, RF) for travel mode choice analysis and addressed the importance of different variables and how they related to different travel modes. Among the investigated classifiers, the RF produced the most accurate predictions, while the performance of MNL was low.

Lindner *et al.* [112] compared the performance of an ANN and a DT with a binary logit in a multicollinear study case for the estimation of motorized travel mode choice. They found that DTs and ANNs can overcome the disadvantages of the more traditional model, especially the constraint of a multicollinear database, and these approaches can also construct a robust non-continuous model from the patterns of the entire sample.

Assi *et al.* [113] compared the effectiveness, robustness, and convergence of extreme learning machine (ELM), SVM, and multi-layer perceptron ANN to predict school-goers mode choice behaviour. Both ELM and ANN outperformed the SVM technique in terms of training and testing accuracies. The SVM technique was more computationally expensive, while the ELM was the best one in terms of computational expense.

Richards and Zill [114] examined several machine learning methods (ANN, RF, Gradient Boosting) to model mode choice decisions and compared the results to a well calibrated nested logit model. All the models performed well regarding both individual level predictive accuracy and the aggregate mode share. The best performing model was gradient boosting with a mean predictive accuracy of 90%.

Zhao *et al.* [115] provided a comparison between machine learning methods (Naïve Bayes, DT, Boosting, Bagging, RF, ANN) and MNL for travel mode choice modelling and evaluated the two approaches on the stated-preference survey data for a new type of transit system. The RF model was the best machine learning model, and it significantly outperformed the MNL both at individual and aggregate levels.

Zhou *et al.* [116] applied several machine learning techniques (stochastic gradient descent, k-nearest neighbour, DT, SVM, Naïve Bayes, AdaBoost, Bagging, RF, Extra Trees, Gradient Boosting, ANN) to simulate the means of transport based on environmental and temporal factors to model travel choices between bike-sharing and taxi. The performance of multi-layer ANN did not surpass that of classical non-linear models (*e.g.*, RF), and within ANNs, a deep ANN only increased the prediction accuracy by a marginal rate.

Aghaabbasi *et al.* [117] investigated the factors that motivate the adoption and the usage frequency of ride-sourcing among students in a public university using RF and Bayesian Network analysis. The predictors with the highest importance from the RF results were included as predictors of the Bayesian Network.

Liang *et al.* [118] presented three methods including a MNL model, a RF and SVM to estimate household travel mode. The accuracies of the three methods stabilized when increasing the sample size, but only up to a certain limit. The accuracies of MNL, RF and SVM were all around 70%, and the results of the RF were consistent with those of the MNL.

Koushik *et al.* [119] reviewed the activity-travel behaviour literature that employs Machine Learning (ML) techniques for empirical analysis and modelling, like SVMs, DTs, ANNs, Bayes Classifiers, and Ensemble Learners. The review found that most of the studies identify the lack of interpretability as a serious shortcoming in ML techniques.

Hillel *et al.* [12] conducted a systematic review of machine learning methodologies for modelling passenger mode choice and identified and quantified the prevalence of methodological limitations in previous studies. The limitations identified in the review highlighted the need for a deeper understanding of the methodologies used for ML modelling of choice behaviour.

Wang *et al.*[120] provided a generalizable empirical benchmark by comparing 105 between machine learning and discrete choice model classifiers from 12 model families (Logit, deep ANNs, discriminant analysis, Bayesian Models, SVMs, K nearest neighbours, DTs, generalized linear models, Gaussian process, RFs, bagging, and boosting) and evaluating both prediction accuracy and computational efficiency of each model. They found that ensemble methods (boosting, bagging, and RFs) and deep ANN achieved the highest predictive performance, but at a relatively high computational cost. Random forests were the most computationally efficient, balancing between prediction and computation.

Table 5. Details for the literature regarding Several Algorithms

Authors	Year	Data	Territorial context
Xie, Lu, and Parkany [109]	2003	**4,746 observations** Two-day travel diary and detailed individual and household sociodemographic data three data sets reflecting household, person, and trip, characteristics.	San Francisco (USA)
Shukla, Ma, Denagamage, and Huynh [110]	2013	**100,000 observations** Household travel survey data including details of each of the trips that each person in a household makes over 24 hours on a day, socio-demographic attributes of households and of individuals.	Sydney (Australia)
Hagenauer and Helbich [111]	2017	**230,608 (100,000) observations** Individual travel diaries in which participants were asked to record every trip over the course of six days, in addition to socio-economic data about the participants as well as information on households.	Netherlands
Lindner, Pitombo, and Cunha [112]	2017	**18,733 observations** Disaggregate data referring to the Origin/ Destination Survey carried out in 2007 by the São Paulo Metropolitan Company.	São Paulo (Brazil)
Assi, Shafiullah, Nahiduzzaman, and Mansoor [113]	2019	**1,484 observations** Questionnaires distributed to students and collected on the next day, designed to gather information about the present mode choice, and trip and household characteristics.	Khobar, Dhahran (Saudi Arabia)
Richards and Zill [114]	2019	**150,000 observations** Dataset comprised of all trips completed on a particular day recorded by selected households.	Melbourne (Australia)
Zhao, Yan, and Van Hentenryck [167]	2019	**8,141 observations** Stated-preference survey completed by the faculty, staff, and students at the University of Michigan.	Michigan (USA)
Zhou, Wang, and Li [116]	2019	**10+ Millions? (30,000) observations** Dataset including both trip data and station information for bike sharing. Dataset including trip information including taxi fare for taxis. Data regarding the built environment.	Chicago (USA)
Aghaabbasi, Shekari, Shah, Olakunle, Armaghani, Moeinaddini [117]	2020	**358 observations** A survey conducted among the students of the Universiti Teknologi Malaysia in the second-largest public university campus in Malaysia.	Skudai (Malaysia)
Liang, Xu, Grant-Muller, and Mussone [118]	2020	**101,053 observations** A survey based on households, which involved vehicle drivers and users of public transport around the city of Milan.	Milan (Italy)
Wang, Mo, Hess, Zhao [120]	2021	**100,000 observations** A national household travel survey collected in the United States, a travel demand survey collected in London, and a stated preference survey collected in in Singapore.	USA; London (UK); Singapore

2.3.5 “Hybrid” Methods

All the works reported in this section are related to studies which tried to find methods to incorporate both classic discrete mode choice modelling and machine learning algorithms, hence why the name “hybrid” methods was chosen for the section.

Gazder and Ratrout [121] investigated the use of Logit-ANN based ensembles in mode choice modelling for different numbers of transportation modes and predictor variables. The results of the proposed method were compared to several other ANN models, and the Logit-ANN models gave better prediction than other models in almost all situations. The ensemble also showed high accuracies for overall as well as individual mode predictions for all multinomial problems.

Sifringer *at al.* [122] brought the predictive strength of neural networks to the field of DCMs by matching the mathematical derivation of the multinomial logit model to its neural network equivalent. They added a term arising from a dense neural network in the utility function, using all discarded features from the original DCM model as input to the NN. This greatly increased the predictive strength of the model, and they highlighted how this NN-derived term fits very well in DCM theory when relating it to a random utility term.

Arkoudi *et al.* [32] proposed an approach that combines theory and data-driven choice models using Neural Networks. They use continuous vector representations (embeddings) for encoding categorical or discrete explanatory variables with a focus on interpretability, by associating each of the embeddings' dimensions to a choice alternative. Their models preserved interpretability of the utility coefficients for all the input variables despite being based on ANN principles, and their results delivered state-of-the-art predictive performance, outperforming existing ANN-based models while drastically reducing the number of parameters.

Sfeir *et al.* [123] presented a Latent Class Choice Model with a flexible class membership, by formulating the latent classes using Gaussian-Bernoulli mixture models. They derived an Expectation-Maximization algorithm is derived for the estimation and compared their model to traditional discrete choice models based on parameter signs, values of time, and goodness-of-fit. Their results showed an improvement in the overall performance of latent class choice models by providing better out-of-sample predication accuracy in addition to better representations of heterogeneity without weakening the behavioural and economic interpretability of the choice models.

Wong and Farooq [33] proposed a Residual Logit (ResLogit) model formulation which integrated a Deep Neural Network architecture into a multinomial logit model. Their approach extended the systematic utility function to incorporate non-linear cross-effects using a series of residual layers. The model structure accounted for cross-effects, choice heterogeneity and other effects in a non-linear manner. Their findings showed that the ResLogit approach significantly outperformed multi-layer perceptron models while providing similar interpretability as a MNL model.

Han *et al.* [34, 124] formulated a model consists of two modules: a neural network (TasteNet) that learns taste parameters as flexible functions of individual characteristics; and a multinomial logit (MNL) model with utility functions defined with expert knowledge. The taste parameters learned by the NN are fed into the choice model and link the two modules. Moreover, they required estimates of behaviour indicators to be realistic at the disaggregated level. They showed that, on the publicly available Swissmetro dataset, their TasteNet-MNL outperformed the predictability of both multinomial logit and Mixed Logit model.

Sfeir *et al.* [35] presented a Gaussian Process – Latent Class Choice Model to integrate a non-parametric class of probabilistic machine learning within discrete choice models with the aim of improving the discrete representations of unobserved heterogeneity. The model would assign individuals to behaviourally homogeneous clusters and simultaneously estimate class-specific choice models by relying on random utility models. Results showed that their approach allows for a more complex and flexible representation of heterogeneity and improves both in-sample fit and out-of-sample predictive power.

Table 6. Details for the literature regarding "Hybrid" Methods

Authors	Year	Data	Territorial context
Gazder and Ratrout [121]	2015	**654 observations** Passenger travel between Khobar-Dammam metropolitan area and Kingdom of Bahrain.	Khobar-Damma, Bahrain (Saudi Arabia)
Sifringer, Lurkin, and Alahi [122]	2018	**10,728 observations** A stated preference survey on transport modes, in which each individual informed of his choice in transportation for various trips including car, train or the innovative Swissmetro project.	Switzerland
Arkoudi, Azevedo, and Pereira [32]	2021	**10,728 + 14,550 observations** A stated preference survey on transport modes from the Swissmetro project, and a subset of the Danish National Travel Survey.	Switzerland; Denmark
Sfeir, Abou-Zeid, Rodrigues, Pereira, and Kaysi [123]	2021	**81,086 + 2,600 observations** A dataset combining trip diaries of the London Travel Demand Survey with alternatives extracted from Google, and a dataset from a stated preferences commuting survey collected in Beirut.	London (UK); Beirut (Lebanon)
Wong and Farooq [33]	2021	**60,365 observations** MTL Trajet dataset collected from the user's smartphone by using a mobile application during a travel survey.	Montreal (Canada)
Han, Pereira, Ben-Akiva, and Zegras [34, 124]	2022	**14,000 + 10,692 observations** A synthetic dataset generated with an underlying logit model, and a stated preference survey on transport modes from the Swissmetro project.	Switzerland
Sfeir, Rodrigues, and Abou-Zeid [35]	2022	**2,600 observations** A dataset from a stated preferences commuting survey collected at the American University of Beirut.	Beirut (Lebanon)

3 USING ARTIFICIAL NEURAL NETWORKS TO ESTIMATE THE VALUE OF TRAVEL TIME IN THE CONTEXT OF MODE CHOICE MODELLING

3.1 Introduction

In our work we propose an alternative method for calculating the value of travel time through the results of machine learning algorithms. We compare different methodologies, so that we could safely say that the soundness of our methodology is validated by the estimation of more classical models and by the use of different databases containing real-life data obtained from travel surveys.

Recently, an increasing number of applications of machine learning (ML) algorithms to choice modelling in transportation have been tested as an alternative to the traditional discrete choice models based on econometric theory. The higher flexibility of ML methods, which generally require no pre-assumptions regarding the mathematical formulation of the underlying relations between the variables playing different roles in explaining a certain phenomenon, is one of the strongest factors affecting this recently developed interest in their possible use in this field of study.

However, ML methods, at least in their purest form, can be considered as black-box methods, which is a far from ideal aspect for an analyst wishing to obtain meaningful information from their outputs. To overcome this inherent limit, some researchers have tried to find alternative ways to use such methods, by either opportunely transforming the outputs of machine learning algorithms so that they can be interpreted from an economic point of view [29–31], or by building modified hybrid versions of them in such a manner that their results can be interpreted in the same way analysts are used to do with classic discrete choice models [32–35]. While obtaining coefficients with their associated statistical significance is practically unfeasible by using machine learning methods, it is instead possible to extrapolate values regarding elasticities and marginal effects of the considered variables, to compare them to those obtained by specifying and estimating econometric models. Some studies have already proven the validity of these methods [30, 95]

One interesting element to analyse is the value of travel time (VTT), since it is a very informative parameter to be considered when analysing the results of choice models in transportation, and when evaluating possible policy implications. In fact, the value that people place on saving total travel time is one of the most important indices that can be inferred [125]. As a matter of fact, since travel generates a derived demand, the time people need to dedicate to their trips is effectively sacrificed, as most people would rather invest this time by doing more desired activities at home, at work or anywhere else. Thus, the time consumed because of the need to travel (walking, waiting or riding a vehicle) normally represents an undesirable circumstance for most individuals, and, therefore, they are willing to exchange another good (in this case, their money) in order to reduce their travel time, and the VTT represent the amount they are willing to pay to reduce their travel time by one unit of time [126]. Understanding the VTT would allow decision-makers to act by introducing policies aimed at reducing the amount of time wasted when making a trip or improving the conditions in which trips are made, and this additional benefits should be accounted among those produced of public investments [127].

To date, the argument of obtaining VTT through ML methods has been studied by few researchers [30, 95], so there is still much to be found out and discussed. The method we propose is independent from the shape and formulation of the mode-choice functions, so it can be applied to both econometric models and ML methods, given the output allows the estimation of individual level mode-choice probabilities. Such a method was necessary because ML methods do not have a mathematical formulation of the utility, hence it not possible to calculate derivatives, needed to calculate the values for elasticities and marginal effects using their correct definition. Instead, we chose to approximate the infinitesimal formulation (which uses the derivatives) with a ratio of finite differences. This method of calculating marginal effects for ML methods, proposed by Zhao *et al.* [31], as far as we know, has not yet been applied to compute the value of travel time.

We applied this method to machine learning methods and then we compared the results with those of some econometric models. In particular, we specified and estimated a multinomial logit and a mixed logit as discrete choice models, while for the machine learning part we constructed a particular specification of a neural network which considers “alternative specific utility functions”, inspired by the work of Wang *et al.* [29].

To test the validity of our methodology we also employed two different datasets. The first dataset contains 7,021 observations, collected through a stated preference survey conducted in Switzerland in 1998 among commuters traveling between St. Gallen and Geneva and considers as a dependent variable the choice to commute by train, Swissmetro (an innovative mag-lev underground system) or car. The second one has instead 2,873 observations and it was collected through a revealed preference survey conducted in 2019-2020 in the metropolitan area of Cagliari (Italy) among a sample of commuters among students, university staff, and public employees. In this case the dependent variable was the choice to commute by using one of the following means of transport: car, public transport, and walking. Both datasets also contain personal and socio-economic characteristics of each individual.

For all the models we split the datasets in a training and testing set, we then estimated the final results using the testing set only. Since neural networks are highly sensitive to the values of their hyper-parameters, we estimated several different models which used different sets of hyper-parameters selected from specific ranges in order to find the optimal calibration values.

Regarding the main outputs, we mainly focused on two dimensions: the first, which was the focus of this analysis, is the value of travel time for both car and public transport alternatives; second, we compared direct- and cross-elasticities for several key level of service variables, in order to verify that our neural network is working correctly and is producing consistent and reliable results.

3.2 Methodology

Generally speaking, the level of utility that a decision maker *n* receives from choosing an alternative *j* can be defined as U_{nj}. However, researchers cannot observe this utility directly, but rather they obtain information regarding one or more aspects $\boldsymbol{x}_{nj}$ which are supposed to influence the utility. Hence, they can only reproduce a representative utility $V_{nj} = f(\boldsymbol{x}_{nj})$, and there exists a relation between the two in the form of $U_{nj} = V_{nj} + \varepsilon_{nj}$. ε_{nj} is a random error component which captures the effects of all the elements which influence U_{nj} but are not included in V_{nj}.

In transportation economic theory, the value of travel time (VTT) is defined as the marginal rate of substitution of the travel cost for the travel time of the same alternative [128], *i.e.* the VTT is equivalent to the marginal utility of travel time MU_{njT} over the marginal utility of the cost MU_{njC}:

$$VTT_{nj} = \frac{MU_{njT}}{MU_{njC}} = \frac{\frac{\partial V_{nj}}{\partial T_{nj}}}{\frac{\partial V_{nj}}{\partial C_{nj}}} \tag{1}$$

where T_{nj} and C_{nj} are respectively the travel time and cost associated to alternative *j* by individual *n*.

However, machine learning algorithms, and neural networks more specifically, do not usually allow to estimate the utilities as straightforwardly as we can with discrete choice models, so it would be also not be possible to obtain a value for the marginal utilities. Since in our NN architecture we obtain probabilities as the output of the model, a relationship between the derivatives of utilities and choice probabilities is needed if we desire to apply relation (1). It is possible to demonstrate [2] that the following relation (2) is true in the case of a multinomial logit:

$$\frac{\partial P_{nj}}{\partial x_{nj}} = \frac{\partial V_{nj}}{\partial x_{nj}} P_{nj}(1 - P_{nj}) \tag{2}$$

If we combine equations 1 and 2, we obtain the relation we were looking for, which directly connects the VTT of individual *n* for alternative *j* to the derivatives of the choice probabilities w.r.t. (with respect to) travel time and cost:

$$\frac{\frac{\partial V_{nj}}{\partial T_{nj}}}{\frac{\partial V_{nj}}{\partial C_{nj}}} = \frac{\frac{\partial V_{nj}}{\partial T_{nj}}}{\frac{\partial V_{nj}}{\partial C_{nj}}} \cdot \frac{P_{nj}\left(1 - P_{nj}\right)}{P_{nj}\left(1 - P_{nj}\right)} = \frac{\frac{\partial V_{nj}}{\partial T_{nj}} P_{nj}\left(1 - P_{nj}\right)}{\frac{\partial V_{nj}}{\partial C_{nj}} P_{nj}\left(1 - P_{nj}\right)} = \frac{\frac{\partial P_{nj}}{\partial T_{nj}}}{\frac{\partial P_{nj}}{\partial C_{nj}}} = VTT_{nj} \tag{3}$$

In order to validate the results obtained through the neural networks, we needed some benchmarks values to verify if our proposed models are working correctly. We chose to use logit models, since they represent the most often used tool to model choice behaviour in transportation. The first model we used is the multinomial logit (MNL) model, which is the simplest and most commonly used among the choice models [2]. Logit models are obtained by assuming that each random error component ε_{nj} is an independently, identically distributed extreme value. In these models, the utility perceived by individual *n* for alternative *j* is specified as a linear combination of observed variables $\boldsymbol{x}_{nj}$ which uses a set of parameters $\boldsymbol{\beta}$ to represent the individuals' tastes:

$$V_{nj} = \boldsymbol{\beta}\boldsymbol{x}_{nj} \tag{4}$$

Then, the choice probability for alternative *j* and individual *n* takes the form of a softmax function [129]:

$$P_{nj} = \frac{e^{V_{nj}}}{\sum_i e^{V_{ni}}} \tag{5}$$

The multinomial logit model satisfies the axiom of the independence of irrelevant alternatives (IIA), meaning that the relative probability of choosing one alternative over another is independent from the other alternatives [2, 50]. The second reference is the mixed logit model (MXL), which is a generalization of the standard logit, for which the utility for individual *n* and alternative *j* is:

$$V_{nj}(\boldsymbol{\omega}) = \boldsymbol{\beta}_{\boldsymbol{\omega}}\boldsymbol{x}_{nj}\,, \qquad \boldsymbol{\beta}_{\boldsymbol{\omega}} \sim f(\boldsymbol{\omega}) \tag{6}$$

where the coefficients $\boldsymbol{\beta}_{\boldsymbol{\omega}}$ follow a distribution $f()$ with parameters $\boldsymbol{\omega}$, and the probabilities of the model are calculated as the integrals of multinomial logit probabilities over the density distribution of the parameters, that can vary randomly across individuals [5]:

$$P_{nj} = \int L_{nj}(\boldsymbol{\omega}) f(\boldsymbol{\omega}) d(\boldsymbol{\omega}), \qquad L_{nj}(\boldsymbol{\omega}) = \frac{e^{V_{nj}(\boldsymbol{\omega})}}{\sum_i e^{V_{ni}(\boldsymbol{\omega})}} \tag{7}$$

The model approach we are proposing is based instead on a neural network architecture. Artificial Neural Networks (NNs) are a family of classifiers which mimic the network structure of the brain [20]. Several versions of NN structures can be found in the literature, but mode choice applications usually rely on the Feed-Forward Neural Network (FFNN) [12]. A FFNN consists of multiple layers of nodes, including an input layer (which passes the values of the attributes to the network), any number of hidden intermediate layers, and an output layer, which returns the predicted values.

NNs can be easily applied to choice analysis, especially if a softmax activation function is applied at the output layer. In this case, the utilities are represented by the output of last intermediate hidden layer, and the value for alternative *j* and individual *n* is:

$$V_{nj} = (g_M^j \circ g_{M-1} \circ \dots \circ g_2 \circ g_1)(\boldsymbol{x}_{nj}) \tag{8}$$

where $\mathrm{g}_k(\boldsymbol{x}_{nj})$ is the transformation applied at the *k*-th hidden layer, and $\mathrm{g}_M^{\mathrm{j}}$ is the one of the last layers.

To build our NN we took inspiration from the study of Wang *et al.* [29], who designed a deep NN architecture with alternative-specific utility functions (ASU-DNN), which improve both the predictive power and interpretability. Their architecture reduces the complexity compared to a fully connected neural network, by stacking *K* subnetworks (where *K* is the number of alternatives), in which each subnetwork receives as its inputs only the alternative-specific attributes and the individual-specific ones. The structure they identified is shown in Figure 1, where it is clear how, after considering only the alternative-specific variables (*X_k*) in the first *M1* layers, they include the individual-level attributes (*Z*) in the remaining *M2* hidden layers. The activation function they used in the hidden layers is obtained by using a linear combination and a rectified linear unit (ReLU) [130].

We modified this architecture to be closer to the model specification of a logit, and because we found out that by making such changes, we found overall better results. First of all, we changed the activation function of the middle layers to a simple linear combination, dropping ReLU altogether since it often led to conflicting results. Second, we decided to use both alternative-specific and individual-level attributes from the beginning, skipping the first *M1* layers used in the ASU-DNN. Third, we differentiated the sub-set of socio-demographic variables based on the alternative, and, most importantly, we included them only in *K-1* alternatives to reproduce the modelling structure used in logits. Finally, we put aside the concept of "deep" NN, since we usually found more consistent results, mainly when observing the signs of the elasticities, when considering a "shallower" architecture, with few layers at maximum. The NN architecture we are proposing is represented in Figure 2. L_0 represents the first hidden layer, which is always present, while L_m represents the set of all M optional hidden layers. Note that, since the NN architecture we are using is based on the structure of a MNL model, we can safely assume that the formulas for the computations of disaggregate indicators (*i.e.*, derivatives and thus elasticities) which can be used for the latter can be considered to be still valid for the former.

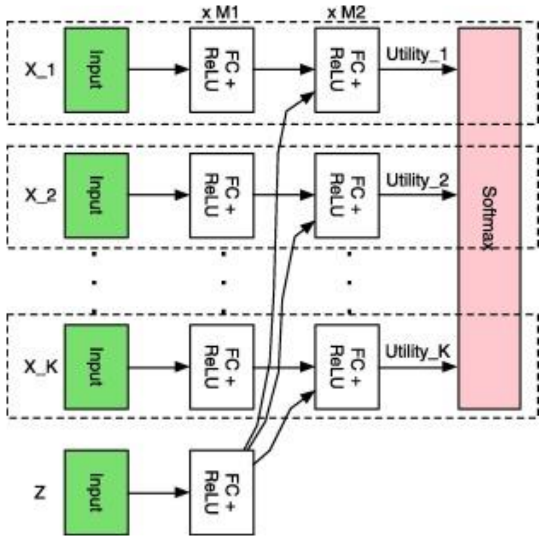


Figure 1. Alternative specific deep NN by Wang et al. [29]

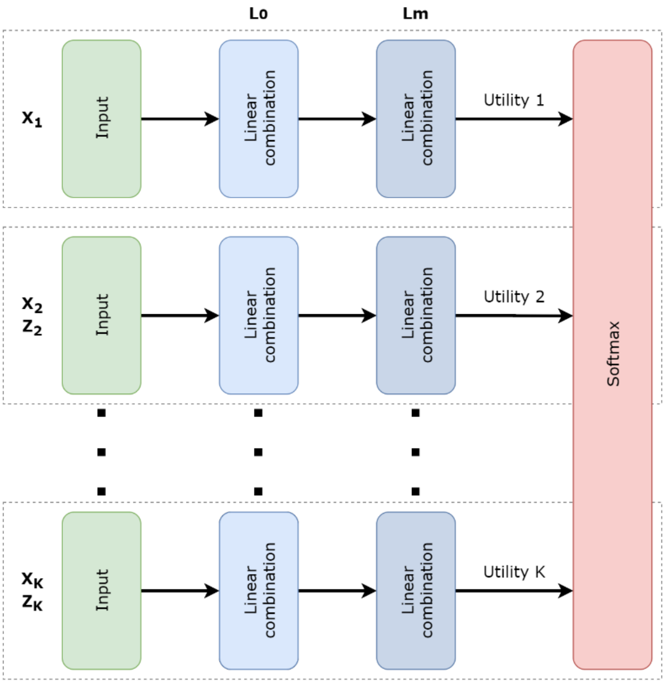


Figure 2. Proposed neural network architecture.

However, since a NN does not provide an analytical form of the probabilities, which would be necessary to calculate the derivatives, we had to approximate the exact formula for this indicator with another quantity which would be comfortably obtainable by using the available outputs, for both logit and NN models. Hence, we moved from the infinitesimal field of the derivatives to a finite difference one, by estimating two different values for the probability, and then using the relation:

$$\frac{\partial P(\boldsymbol{x}_n)}{\partial Z_n} \cong \frac{P(\boldsymbol{x}_n + \Delta Z_n)_j - P(\boldsymbol{x}_n)_j}{\Delta Z_n} \tag{9}$$

in which $P(\boldsymbol{x}_n)_j$ is the probability obtained with the original data, while $P(\boldsymbol{x}_n + \Delta Z_n)_j$ is the probability in the case where the variable with respect to which we wanted to derive (Z_n) was increased by a given factor $q > 0$, *i.e.* $Z_n^* = Z_n + qZ_n = Z_n + \Delta Z_n$. The same reasoning was used to obtain values for the elasticities, since they would also need the computation of derivatives of the probability. In our case, we chose the value $q = 0.01$, so that elasticities could be correctly interpreted as the variation in the percentage of the choice probability when the considered variable increases by 1%. Table 7 shows a comparison of the approximate finite difference formulation compared to the correct one, for both elasticities and value of travel time. Notice that in this case a single value is obtained by averaging the values obtained from all individuals.

Table 7. Comparison between correct formulas and discrete approximations for elasticities and VTT

Indicator	Correct formulation	Finite difference approximation
Elasticity (variable Z, alternative *j*)	$\frac{1}{N}\sum_{n=1}^{N}\left\{\frac{\partial P(\boldsymbol{x}_n)_{ij}}{\partial Z_n} \cdot \frac{Z_n}{P(\boldsymbol{x}_n)_{ij}}\right\}$	$\frac{1}{N}\sum_{n=1}^{N}\left\{\frac{P(\boldsymbol{x}_n + \Delta Z_n)_j - P(\boldsymbol{x}_n)_j}{\Delta Z_n} \cdot \frac{Z_n}{P(\boldsymbol{x}_n)_j}\right\}$
Value of Travel Time (alternative *j*)	$\frac{1}{N}\sum_{n=1}^{N}\left\{\frac{\frac{\partial P(\boldsymbol{x}_n)_j}{\partial T_{nj}}}{\frac{\partial P(\boldsymbol{x}_n)_j}{\partial C_{nj}}}\right\}$	$\frac{1}{N}\sum_{n=1}^{N}\left\{\frac{\frac{P(\boldsymbol{x}_n + \Delta T_{nj})_j - P(\boldsymbol{x}_n)_j}{\Delta T_{nj}}}{\frac{P(\boldsymbol{x}_n + \Delta C_{nj})_j - P(\boldsymbol{x}_n)_j}{\Delta C_{nj}}}\right\}$

3.3 Data analysis

3.3.1 Swissmetro

The first dataset we used to estimate the results of our models is the *Swissmetro* dataset. The *Swissmetro* data comes from a two-stage interview which intercepted 442 respondents approached on the train between St. Gallen and Geneva during March 1998. An initial interview was needed to get information about the trip, and it was followed by an SP experiment based on that trip. Nine stated-choice situations were generated for each respondent, offering as alternatives rail, Swissmetro and car.

Each alternative was described by their travel time, cost and headway if relevant. Following that, several relevant licence plates were recorded during September 1997 at Zürich, Bern and Geneva, and the owners of these cars were sent a survey-pack, which asked them about the trip and enquired about their willingness to participate in a second survey. A total of 770 people completed the survey and participated in a SP survey, which was generated using the same approach which had been used for the rail interviews. For more details on the data collection procedure, see the work from Bierlaire *et al.* [131].

From the total sample of 10,728 observations, we used some filters to obtain a sub-sample more suitable for our purpose. We removed all those individuals for which age and/or income were not known, we excluded those which had a yearly subscription (since their cost for train and Swissmetro was expressed as the annual total and would have led to an overestimation of the VTT), and finally we removed some outliers based on the ratio of travel time and cost. Our final dataset contains 7,021 observations, and Table 8 shows the analysis of the data therein contained. Regarding mode choice, almost 58% chose the new alternative, 36.5% the car and the remaining the train. All observations had all modes available, 46% travelled without luggage, 53% had a single item, and less than 1% had two or more.

Table 8. Data analysis – Swissmetro database

Categorical variables

		N	%			N	%
Total sample		7,021		Age	≤ 24 y.o.	126	1.79%
Choice	Train	404	5.75%		> 24 - ≤ 39 y.o.	2,029	28.90%
	Swissmetro	4,056	57.77%		> 39 - ≤ 54 y.o.	2,848	40.56%
	Car	2,561	36.48%		> 54 - ≤ 65 y.o.	1,532	21.82%
Availability	Train	7,021	100%		> 65 y.o.	486	6.92%
	Swissmetro	7,021	100%	Income	< 50,000	993	14.14%
	Car	7,021	100%	[CHF/year]	50,000 - 100,000	2,921	41.60%
Luggage	None	3,242	46.18%		> 100,000	3,107	44.25%
	One	3,727	53.08%	Gender	Male	5,719	81.46%
	Several	52	0.74%		Female	1,302	18.54%

Continuous variables

		Avg.	**St. Dev.**
Travel time [min]	Train	184.38	78.6286
	Swissmetro	97.16	58.3361
	Car	154.67	78.7397
Cost [CHF]	Train	89.27	50.4818
	Swissmetro	107.56	61.0139
	Car	96.90	45.8759
Headway [min]	Train	70.17	37.4871
	Swissmetro	20.09	8.1923

Most of the respondents (91.3%) are between 24 and 65 years old, with roughly 40% of the sample aged between 39 and 54. The majority of the respondents (86%) declared an income of 50,000 CHF/year or more, with 44% declaring incomes higher than 100,000 CHF/year, while just 14% of them earned under 50,000 CHF/year. The sample is heavily skewed towards the male population (81.5%), since only 18.5% of the respondents were female. The average travel time with train shows how it is the slower of the three alternatives, with slightly more than 3 hours per trip, Swissmetro was the faster with 97 minutes, while car was in between (154 minutes). In contrast, Swissmetro was also the most expensive choice, with an average of 107.56 CHF/trip, followed by car (97 CHF/trip) and then train (90 CHF/trip), even though the differences are less noticeable when compared to travel time. Finally, Swissmetro also provides lower waiting times, with an average headway of 20 minutes, compared to the 70 minutes of the train alternative.

3.3.2 Svolta Cagliari

The second dataset we used is the *Svolta Cagliari* dataset, which was collected starting from 2019 during an experimental programme lead by the government of the city of Cagliari (Italy), in collaboration with the University of Cagliari. The aim of the program was that of intercepting commuters which frequently visit the city of Cagliari, mainly for working or studying, to find possible ways of convincing them to change their travel behaviour towards more sustainable means of transport. An online survey was distributed either by means of direct contact via e-mail, or by placing advertisements, both physically (with billboards spread in the city and onboard public transport vehicles, and posters placed in buildings frequented by university students and staff, and offices used by public administration employees) and digitally (banners on institutional websites of public administrations and public transport operators, on websites of the local news outlets, and social media). The original text of the survey (in Italian) is shown in Appendix A. The complete questionnaires were over 4,000, but we analysed a sub-sample of the people which travelled by either car, public transport (PT) or walking, and had at least two of these alternative available. The availability for each alternative were defined as:

- *car*: the individuals possess a driver licence, and they have access to at least one car in their household;
- *public transport*: the distance from the bus stop /station in 2 km or less from the individuals' home, and the headway associated with the transport service is 30 minutes at maximum;
- *walking*: the distance on foot must be 5 km or less, and the path must not use any extra-urban road or infrastructure.

The final dataset, described in Table 9, is composed by 2,873 observations. Modal choice is split in 52% car, 38% public transport and 10% walking, and availability of car and PT for work-trips is very high, 86% and 94% respectively, while only 44% of the respondents could walk to their workplace. Almost everyone possesses a driving licence (96%) and most of them also own a car (86%), but less than half the respondents own a bike (49%). Regarding where their residence is located, most of them (62%) live inside the city of Cagliari, the rest lives either in the surrounding metropolitan area (28%) or somewhere else on the island (10%). Most of the respondents fall in the age ranges of 18 to 30 years old and 40 to 60 years old, which are represented by almost equivalent portions of the sample (39.6%), while 30 to 40 years old are 15% and just over 6% are over 60 years old. The gender ratio is almost evenly split, with a slightly higher percentage of females (54.5%) compared to males (45.5%). The majority of the respondents were public employees (50%), 41% were university students (working towards a degree or a Ph.D.), 5.5% were employers, and the remaining were composed by people which were unemployed or retired. 45% of the respondents possess at least a high-school diploma, and 51% have a university education, with 16.5% having a Ph.D.

A considerable portion of the sample (38%) declared earning 1,000 €/month or less (to be expected, considering the number of students), 45% had an income between 1,000 and 2,000 €/month, and 18% earned more than 2,000 €/month. When analysing the composition of the household, we observed an average number of members of around 3, less than half of the families have children, and even less under 10 years old. On average, each household possesses 1.8 cars.

The average number of work / study related trips across the sample is 220 per year. Most of the respondents (83.5%) declared they did not stop during the trip they described, while 16.5% stopped at least once along the route. 44% of the respondents left their home in the 7:30-8:29 AM time slot, followed by the 8:30-9:29 AM one with 29%. 12.4% started their trip early in the morning (earlier than 7:30 AM), and 8% between 9:30 and 12:29 PM, while the remaining 7% departed later in the day.

The analysis of the level of service variables, for each alternative, is made with reference to only those individuals which have that alternative available. On average, the proposed car trips lasted 16.1 minutes, with a generalized cost of 1.80 €/trip. For those who actually chose to travel by car, the duration was 15.9 minutes instead, and the cost 1.93 €/trip, while for those who did not use cars, 16.3 minutes at 2.14 €/trip. The walking trips averaged a duration of 29.0 minutes, but times were definitely shorter for those who actually walked (18.5 minutes) and longer for those who did not (32.1 minutes).

Table 9. Data analysis – Svolta Cagliari database

Categorical variables

		N	%			N	%
Total sample		2,873		Age	≥ 18 - ≤ 30 y.o.	1,138	39.61%
Choice	Car	1,484	51.65%		> 30 - ≤ 40 y.o.	418	14.55%
	Public transport	1,093	38.04%		> 40 - ≤ 60 y.o.	1,136	39.54%
	Walking	296	10.30%		> 60 y.o.	181	6.30%
Availability	Car	2,470	85.97%	Gender	Male	1,308	45.53%
	Public transport	2,716	94.54%		Female	1,565	54.47%
	Walking	1,272	44.27%	Occupation	Employee	1,433	49.88%
Driving licence		2,758	96.00%		Student	1,043	36.30%
Owns a car		2,458	85.56%		Employer	157	5.46%
Owns a bicycle		1,409	49.04%		Ph.D.	131	4.56%
Departure time	5:30 - 7:29 AM	357	12.43%		Unemployed	42	1.46%
	7:30 - 8:29 AM	1,255	43.68%		Retired	35	1.22%
	8:30 - 9:29 AM	835	29.06%		Student-worker	24	0.84%
	9:30 AM - 12:29 PM	229	7.97%		Homemaker	8	0.28%
	12:30 - 14:29 PM	54	1.88%	Education	Up to middle school	69	2.40%
	14:30 - 16:59 PM	94	3.27%		High school	1,283	44.66%
	17:00 - 19:29 PM	37	1.29%		Specialization	52	1.81%
	19:30 - 22:00 PM	12	0.42%		Degree	994	34.60%
Trip stops	Yes	475	16.53%		Ph.D.	475	16.53%
	No	2,398	83.47%	Income	≤ 500	799	27.81%
House location	Cagliari city	1,790	62.30%	[€/month]	> 500 - ≤ 1,000	284	9.89%
	Metropolitan area	806	28.05%		> 1,000 - ≤ 1,500	661	23.01%
	South Sardinia	260	9.05%		> 1,500 - ≤ 2,000	623	21.68%
	Other	17	0.59%		> 2,000 - ≤ 3,000	294	10.23%
					> 3,000	212	7.38%

Continuous variables

		Avg.	St. Dev.		Avg.	St. Dev.
Household	Nr. of members	3.08	1.2729	Trips / year	220.52	76.6193
	Nr. of children	0.48	0.8455			
	Nr. of children < 10y.o.	0.18	0.4819			
	Nr. of cars	1.79	0.8206			

Level of service variables

		Alternative is available		Alternative is chosen		Alternative is not chosen (if available)	
		Avg.	St. Dev.	Avg.	St. Dev.	Avg.	St. Dev.
Car	Travel time [min]	16.07	11.6289	15.91	9.8216	16.32	13.9154
	Cost [€]	2.01	2.5300	1.93	2.0985	2.14	3.0637
Public transport	Travel time [min]	23.94	20.1651	24.12	22.9273	23.83	18.0751
	Cost [€]	1.04	0.8876	1.08	1.0445	1.02	0.7633
	Walk time [min]	9.29	5.4052	8.90	5.2243	9.56	5.5093
	Wait time [min]	3.81	1.4479	3.82	1.4083	3.80	1.4745
	Headway [min]	2.94	5.4068	2.64	6.2397	3.14	4.7558
	Transfers	0.40	0.6032	0.33	0.5966	0.44	0.6034
Walking	Travel time [min]	28.95	13.7168	18.49	9.8892	32.12	13.1288

Public transport trips lasted 23.9 minutes on average (in-vehicle time), with a cost of 1.04 €/trip, and these values are basically the same whether the individuals actually used PT (24.1 minutes at the cost of 1.08 €/trip) or if they used other means of transport (23.8 minutes and 1.02 €/trip). It is actually interesting to observe how the average values are slightly higher for those who used public transport for their trips. The average walking time to get to the closest bus stop/station is 9.3 minutes (8.9 minutes if PT was the chosen alternative, 9.6 minutes otherwise), the waiting time is 3.8 minutes (in any case) and the headway is 2.9 minutes (2.6 minutes if chose, 3.1 otherwise). The average number of transfers for each trip is 0.4, meaning that less than half of them included a change of vehicle/line (0.3 for those who actually chose public transport, 0.4 for the others).

3.4 Model specification

3.4.1 Variable selection

To test all the models equally, the same set of variables had to be used on all of them. This however required a reliable method to select the variables we wanted to use. Since logit models allow to recognize which attributes can be more useful than others, by means of statistical significance (t-stats and p-values), while also being sensible to the set of variables used for estimation, we chose to select the variables based on the results of a multinomial logit, in which variables were removed one by one until all of their associated parameters were statistically significant. This in turn allowed us to get the results from logit models in the best possible conditions, thus giving them an "advantage" when comparing the values of the outputs. This also means that, if the NN results are comparable to the ones obtained from the logits, then we can safely assume they stand on equal footing to a MNL estimated in optimal conditions. We applied the same method to both datasets, and the selected variables are shown in Table 10 and Table 11, along with the robust t-test values associated with the corresponding parameters. All these results were obtained by using PythonBiogeme [132].

Table 10. Variables used for the Swissmetro dataset with their MNL parameters and their statistical significance

Variable name	Value	Robust t-test	Variable name	Value	Robust t-test
Train attributes			*Swissmetro attributes*		
Time	-0.013	-8.41	ASC	0.145	0.56
Cost	-0.025	-9.94	Time	-0.011	-7.58
Headway	-0.008	-4.86	Cost	-0.016	-18.30
Car attributes			Headway	-0.013	-3.69
ASC	-0.413	-1.62	Seats	0.808	5.51
Time	-0.015	-9.06	Luggage = 1	-0.288	-4.70
Cost	-0.010	-6.60	Luggage > 1	-1.160	-2.50
Luggage > 1	-1.440	-2.82	Gender = Male	0.647	4.50
Gender = Male	0.659	4.48	Age >39 / ≤54	-0.485	-2.99
Age >39 / ≤54	-0.425	-2.55	Age >54 / ≤65	-1.110	-6.50
Age >54 / ≤65	-0.868	-4.88	Age >65	-1.330	-6.03
Age >65	-0.664	-3.02	Income >50k / ≤100k	0.412	2.55
Income >50k / ≤100k	0.476	2.86	Income >100k	0.851	4.63
Income >100k	0.634	3.34			

Table 11. Variables used for the Svolta Cagliari dataset with their MNL parameters and their statistical significance

Variable name	Value	Robust t-test	Variable name	Value	Robust t-test
PT attributes			*Car attributes*		
ASC	2.510	5.92	Time	-0.046	-2.44
Time	-0.029	-4.75	Cost	-0.162	-1.87
Cost	-0.207	-1.81	*Walking attributes*		
Dep. time 8:30-9:29	-0.310	-2.50	ASC	3.990	8.85
Stop = yes	0.892	5.75	Time	-0.137	-11.47
Age >30 / ≤40	-0.711	-3.82	Owns a car = yes	-2.150	-5.82
Age >40 / ≤60	-0.365	-2.05			
Gender = Male	-0.390	-2.78			
Occup. = Employee	-0.956	-6.18			
Nr. of HH children	-0.708	-6.45			
Nr. of HH members	0.220	3.86			
Owns a car = yes	-2.500	-7.88			

3.4.2 Hyperparameter calibration

NN models are not identifiable, because the empirical risk minimization (ERM) is non-convex with high dimensionality, so their training is very sensitive to the initialization [133, 134]. This also means that, with different initializations, a NN model can end at a local minimum rather than at the global optimum [135, 136]. This issue does not arise in classical MNL models, because their ERM models is globally convex [137]. This non-identification problem ultimately means that each training of a specific NN can lead to very different results, even when considering the same hyperparameters and training sample. To try and reduce this issue, we calibrated the hyperparameters of the NN models by trained the same model and estimate the results several times, systematically changing one of the hyperparameters while keeping the others fixed. In particular, we focused on the number of epochs to be used, the number M of additional hidden layers, and the number of hidden nodes to be used in each of the subnetworks (see section 4.2 Methodology). Table 12 shows the definitive values considered during the estimation of the final results. For additional information on the results that we analysed to get to these values, refer to Appendix B. It is worth of notice how, as we hinted earlier, the number of additional hidden layer is zero, meaning we only have the initial L_0 hidden layer, and so we found the best results by not using a "deep" NN architecture. The neural network models were estimated using the PyTorch library for Python [138].

Table 12. Hyperparameters sets for both datasets

DATASET	N. epochs	N. additional hidden layers	N. hidden nodes
Swissmetro	200	0	30
Svolta Cagliari	1,000	0	50

3.5 Results

In the following we show the results we obtained by training a total of 100 NN models for each dataset, using the hyperparameters in Table 12, and a training set which included 80% of the original dataset. The results were then obtained by predicting the desired indicators (log-likelihood, elasticities, and value of travel time) using the trained model and the testing set (20% of the data). While obtaining the value of travel time was the main goal of this study, we also decide to show some of the direct elasticities, to demonstrate how our model is working correctly in every aspect and not only when estimating some of the possible indicators.

3.5.1 Swissmetro

Table 13 shows a comparison for all the direct- and cross-elasticities obtained by observing the change in probability caused by the variation of alternative-specific level of service variables we considered from the Swissmetro database. Since we wanted to use these indicators as a benchmark, to check the correct operation of the models, we limited ourselves to these elasticities because the microeconomic theory can help us recognize the correct signs (negative for direct and positive for cross-elasticities), and we thus overlooked the one we could have obtained by also using the socio-economic variables, whose interpretation is less direct. We compared the values we extracted from the neural network models (NN) for all indicators, with those obtained through a multinomial logit model (MNL) and a mixed logit model (MXL) estimated with the same set of variables. The first notable result is that all signs are consistent with the microeconomic theory, *i.e.,* all direct elasticities are negative, while all cross-elasticities are positive. Regarding the single values, we will analyse them based on the corresponding variable:

- *Train travel time:* we can see how the direct elasticity produced basically the same value for NN and MNL (-2.30) and a lower absolute value for MXL (-1.39), while cross elasticities values are very close for all models (~0.11-0.12).
- *Train cost*: the pattern is basically the same we observed for *Train travel time*, since we found very close values for the direct elasticities for NN (-2.25) and MNL (-2.22), and a lower effect for the MXL (-1.74), while cross-elasticities showed similar values across all models (~0.08-0.09).
- *Swissmetro travel time*: in this case, direct elasticities produced quite similar values for all models, but MXL results (-0.58) are still distinct from the ones we had with NN and MXL (-0.52 and -0.49 respectively); however, this time we observed more variability in the cross-elasticities, since, although the ones concerning the *train* alternative were very close to each other (~0.58-0.59), the ones relative to *car* obtain through the MXL were higher (0.90) compared to the ones we got from NN and MNL (0.59).

- *Swissmetro cost*: for this variable, while direct elasticities were quite close for all models, the values are all distinguishable, with -0.87 for the NN -0.81 for the MNL, and -0.95 for the MXL; for this attribute, we also had some variability in the cross-elasticities, with the ones of MNL and NN still close to each other (0.97-0.98 for both *train* and *car*), while those for MXL were lower for *train* (0.81) and higher for *car* (1.11).
- *Car travel time*: in this case, we noticed a clear difference for the fact that now direct elasticities were similar for both logits (-1.56 for MNL and -1.57 for MXL) and slightly higher in absolute value for the NN (-1.69); cross-elasticities are also distinct, with 0.89 for NN, 0.78 for MNL, and 0.61 (*train*) and 1.07 (*Swissmetro*) for MXL.
- *Car cost*: lastly, this variable showed a pattern similar to the one we saw for *car travel time*, with elasticities in general being different for all models; direct elasticities are -0.55 (NN), 0.65 (MNL) and -0.61 (MXL), cross-elasticities are 0.31 (NN), 0.34 (MNL) and 0.21-0.33 (MXL, for *train* and *Swissmetro* respectively).

Table 13. Elasticities comparison of MNL, MXL, NN for the Swissmetro dataset

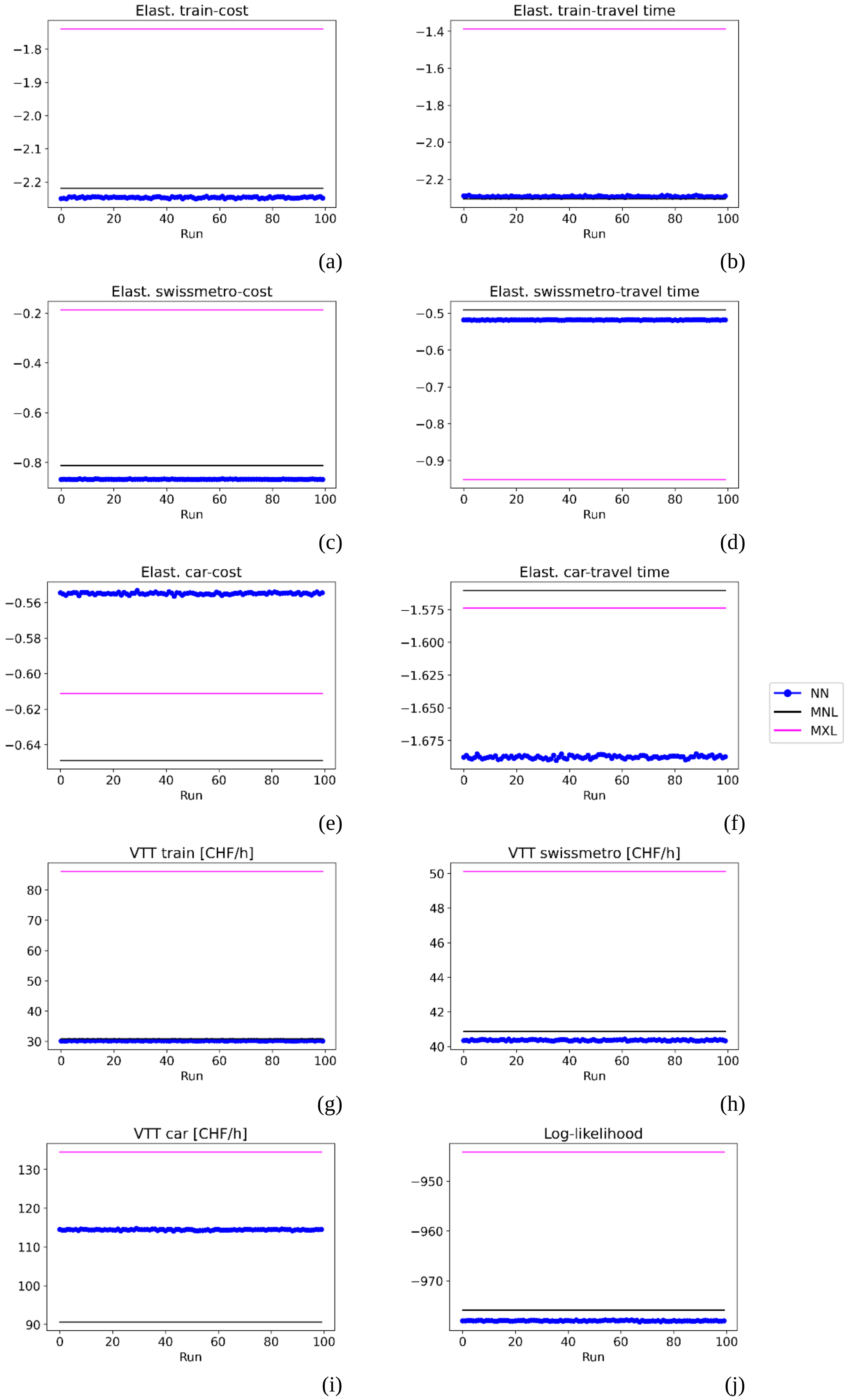
Variable	Model	Train	Swissmetro	Car
Train travel time	**MNL**	-2.3063	0.1098	0.1098
	MXL	-1.3879	0.1225	0.1102
	NN	-2.2934	0.1129	0.1129
Train cost	**MNL**	-2.2187	0.0918	0.0918
	MXL	-1.7393	0.0860	0.0759
	NN	-2.2466	0.0959	0.0959
Swissmetro travel time	**MNL**	0.5970	-0.4915	0.5970
	MXL	0.5802	-0.5893	0.9049
	NN	0.5961	-0.5186	0.5960
Swissmetro cost	**MNL**	0.9731	-0.8125	0.9731
	MXL	0.8132	-0.9517	1.1100
	NN	0.9837	-0.8675	0.9837
Car travel time	**MNL**	0.7797	0.7797	-1.5605
	MXL	0.6114	1.0759	-1.5740
	NN	0.8922	0.8926	-1.6876
Car cost	**MNL**	0.3380	0.3380	-0.6489
	MXL	0.2071	0.3316	-0.6113
	NN	0.3087	0.3088	-0.5549

Figure 3 shows the main results we obtained with all the 100 models trained with the Swissmetro dataset. In order to make sure that different initializations were not an issue (see Paragraph 3.4.2), we mainly focused on the distribution produced by the different indicators, to see if there were any noticeable differences between the several runs. However, for the Swissmetro database at least, we could not observe any relevant phenomenon, since for all the direct elasticities and VTTs the distribution was practically uniform. As a comparison, we also represented in the same graphs the value obtained with MNL and MXL. Overall, the analysis of the elasticities demonstrates how the NN we are using can replicate the behaviour of a logit model, especially an MNL. This further consolidates our belief that a NN model can be successfully used to extract econometric indicators, at least when using this particular database.

Considering all the results, using our neural network we obtained very similar values to those of a multinomial logit (the only exception seems to be the VTT for the *car* alternative), while in general there were noticeable differences with the mixed logit. This means that, as concerns the Swissmetro dataset, the performance of the NN we used confirms the fact that this model can be safely used as an alternative to a MNL, especially if the main interest of the analysis is that of observing a series of econometric indicators.

Table 14. Values of travel time and log-likelihood comparison of MNL, MXL, NN for the Swissmetro dataset

Model	**Log-likelihood**	**VTT Train [CHF/h]**	**VTT Swissmetro [CHF/h]**	**VTT Car [CHF/h]**
MNL	-975.8644	30.7799	40.8705	90.6126
MXL	-944.1914	86.0602	50.1155	134.4626
NN	-978.0179	30.1923	40.3676	114.4333



(a) elasticity of train choice probability w.r.t. train cost; (b) elasticity of train choice probability w.r.t. train travel time; (c) elasticity of Swissmetro choice probability w.r.t. Swissmetro cost; (d) elasticity of Swissmetro choice probability w.r.t. Swissmetro travel time; (e) elasticity of car choice probability w.r.t. car cost; (f) elasticity of car choice probability w.r.t. car travel time; (g) value of travel time for the train alternative; (h) value of travel time for the Swissmetro alternative; (i) value of travel time for the car alternative; (j) log-likelihood

Figure 3. Neural network results for the Swissmetro Dataset

3.5.2 Svolta Cagliari

To further confirm the results we obtained with the *Swissmetro* dataset, we had to use a different source of data to try and obtain similar outcomes. The *Svolta Cagliari* dataset is quite different from the former, since it was collected in a different territorial context (southern Italy *vs.* Switzerland) and later in time (2019 *vs.* 1998), a different survey methodology was used (revealed preferences vs. stated preferences), and finally it is relatively smaller in size (2,873 *vs.* 7,021 observations).

Table 15 shows the comparison for all the direct- and cross-elasticities obtained by observing the change in probability caused by the variation of alternative-specific level of service variables from the *Svolta Cagliari* dataset. We again compared the values obtained with the NN models for all indicators, with those we got from a MNL and a MXL estimated with the same set of variables. We notice also in this case how all the signs we obtained are consistent with the microeconomic theory, *i.e.*, all direct elasticities are negative, while all cross-elasticities are positive. Like we did in the previous case values, we are going to analyse them based on the corresponding variable:

- *Car travel time:* we observed some variability in the elasticities we computed with the three models, since the direct elasticities are lowest absolute value (-0.23) for the NN, followed by the MNL (-0.28) and finally highest (-0.34) for MXL; the cross elasticities follow a similar pattern, with those for the mode *PT* being 0.34 (NN), 0.45 (MNL), and 0.57 (MXL), and those for walking 0.14 (NN), 0.23 (MNL), and 0.34 (MXL).
- *Car cost*: the elasticity values we obtained are all similar among the three models, the direct ones are all between -0.12 and -0.13, the cross-elasticities for *PT* are 0.17 (NN), 0.19 (MNL), and 0.21 (MXL), while those for *walking* are 0.04 (NN), 0.06 (MNL), and 0.07 (MXL).
- *PT travel time*: in this case we observed a pattern similar to the one we had for *Car travel time*, with the NN producing the lowest absolute values (-0.65 for the direct elasticity, 0.37 and 0.26 for the cross-elasticities for *car* and *walking* respectively), MNL in the middle (in the same order, -0.68, 0.43, and 0.32), and MXL returning the highest elasticities (-0.76, 0.53, and 0.50).
- *PT cost*: this variable lead to a pattern in the elasticities similar to those we saw in many cases for the *Swissmetro* dataset, since we obtained very close values with both NN and MNL, and slightly higher ones with MXL; more specifically, direct elasticities are -0.13 for NN/MNL and -0.19 for MXL, cross-elasticities for *car* 0.08 for NN/MNL and 0.11 for MXL, and cross-elasticities for *walking* are -0.07 for NN/MNL and 0.11 for MXL.
- *Walking travel time*: in this final case, direct elasticities are quite close for all the models, -3.41 for NN, -3.52 for MNL, and -3.72 for MXL; cross elasticities are instead very different for the NN, since they

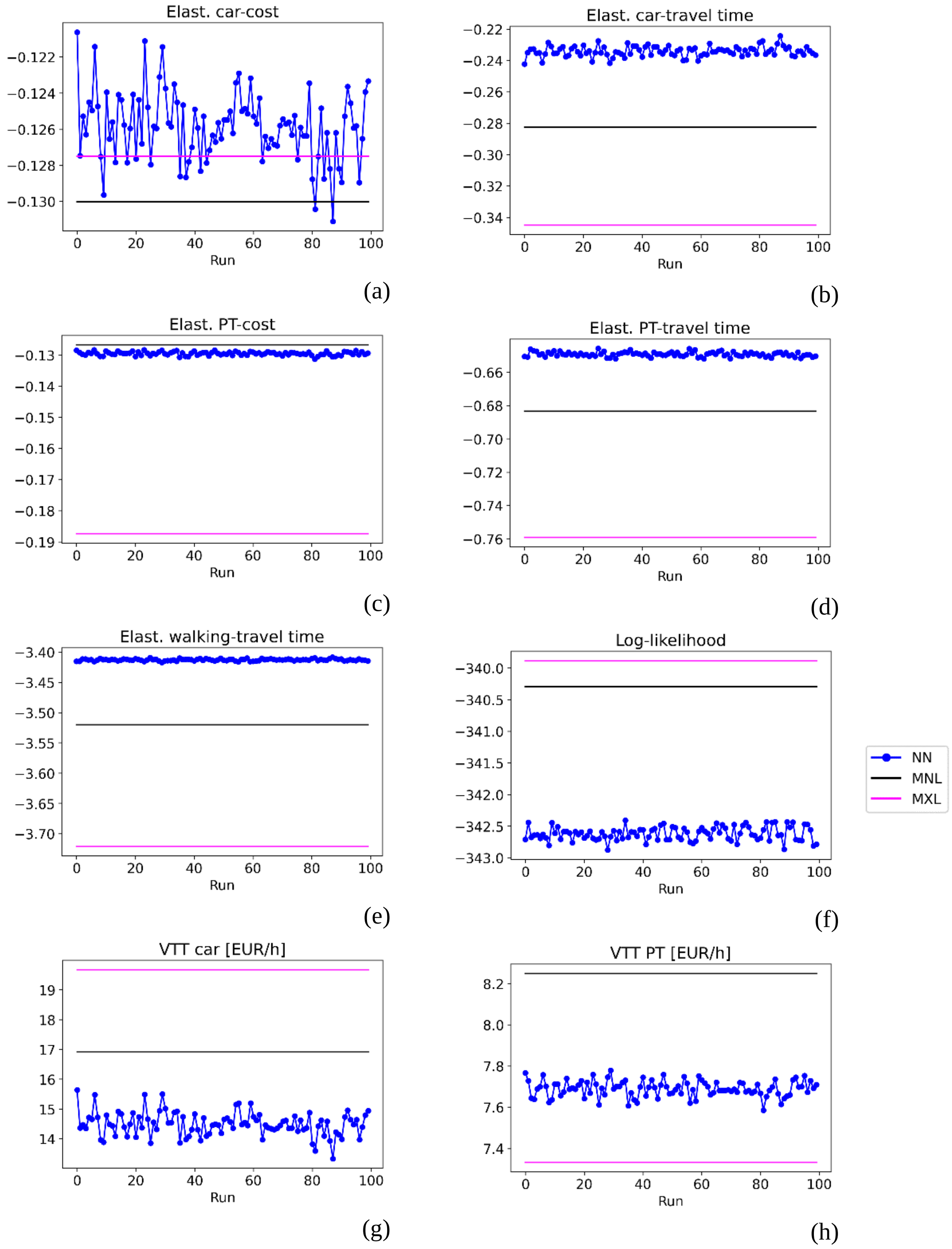
are 0.22 for *car* (compared to 0.50-0.53 for the logits) and 0.29 for *PT* (vs. 0.56-0.59 for MNL and MXL).

Figure 4 shows the main results we obtained with the 100 models trained with the *Svolta Cagliari* dataset. Like we previously said for the *Swissmetro* dataset, we focused on the distribution produced by the different indicators, to see if there were any differences between the several different initializations. Also in this case, we represented the value obtained with MNL and MXL in the same graph, and ultimately the distribution was practically uniform for all the direct elasticities and VTTs. Like we found in the case of the *Swissmetro* dataset, the distributions for all the direct elasticities and the VTTs do not drift too much from the average values. Also in this case, we compared the NN results graphically with those obtained with the MNL and MXL. These results further confirm the fact that NNs ca be employed to estimate econometric indicators.

Finally, Table 16 shows the log-likelihood and value of travel time results for all three models. While the NN showed again the worst performance in terms of fit, with a log-likelihood of -342.62, the other two models ended up with a value not too far off (-340.29 and -339.89), and, ultimately, we can consider them equally standing. Values of travel time for *car* are quite similar, NN predicted the lowest value at 14.50 €/h, then we have MNL with 16.92 €/h, and the highest value was the one given by MXL, 19.67 €/h. The VTT for *PT* showed even closer values, even though here the lowest value was the one from MXL (7.33 €/h), followed by the NN (7.69 €/h) and finally by MNL (8.25 €/h).

Table 15. Elasticities comparison of MNL, MXL, NN for the Svolta Cagliari dataset

Variable	Model	Car	PT	Walking
Car travel time	**MNL**	-0.2826	0.4467	0.2328
	MXL	-0.3449	0.5710	0.3453
	NN	-0.2342	0.3390	0.1411
Car cost	**MNL**	-0.1300	0.1952	0.0564
	MXL	-0.1275	0.2073	0.0722
	NN	-0.1259	0.1732	0.0398
PT travel time	**MNL**	0.4355	-0.6834	0.3171
	MXL	0.5329	-0.7591	0.5059
	NN	0.3691	-0.6492	0.2650
PT cost	**MNL**	0.0855	-0.1268	0.0720
	MXL	0.1138	-0.1874	0.1180
	NN	0.0777	-0.1295	0.0644
Walking travel time	**MNL**	0.5003	0.5621	-3.5203
	MXL	0.5336	0.5884	-3.7208
	NN	0.2216	0.2907	-3.4128



(f)

(a) elasticity of car choice probability w.r.t. car cost; (b) elasticity of car choice probability w.r.t. car travel time; (c) elasticity of PT choice probability w.r.t. PT cost; (d) elasticity of PT choice probability w.r.t. PT travel time; (e) elasticity of walking choice probability w.r.t. walking travel time; (f) log-likelihood; (g) value of travel time for the car alternative; (h)value of travel time for the PT alternative

Figure 4. Neural network results for the Svolta Cagliari Dataset

Table 16. Values of travel time and log-likelihood comparison of MNL, MXL, NN for the Svolta Cagliari dataset

Model	Log-likelihood	VTT Car	VTT PT
MNL	-340.2924	16.9176	8.2491
MXL	-339.8868	19.6699	7.3316
NN	-342.6181	14.4973	7.6898

Once again, even if in this case there are some differences, we can certainly state that our NN model is able to mimic the behaviour of a MNL model. Accounting for all these results, other than the ones we previously obtained with *Swissmetro*, and considering again the major differences in the two datasets we previously listed, we can safely assume that this neural network is a valid alternative to a multinomial logit, especially when the main target of the analysis is that of observing econometric indicators, like elasticities and value of travel time.

4 A NOVEL APPLICATION OF PSYCHO-ATTITUDINAL VARIABLES TO NEURAL NETWORK FOR CHOICE MODELLING IN TRANSPORTATION

4.1 Introduction

Traditional discrete choice models, such as multinomial logit and probit, focus only on directly observable variables, like the attributes of the alternatives and the socioeconomic characteristics of the individuals [6]. However results from recent studies in social sciences have shown that choice behaviour can also be influenced by factors like attitudes, perceptions, norms, intentions, habits, *etc.* [7]. More recent formulations of DCMs include, among their explanatory attributes, latent constructs for capturing the impact of such subjective factors. These are usually called hybrid choice models or integrated choice and latent variable (ICLV) models [8–10]. This class of models have recently gained exposure because of an increasing interest in including latent constructs to capture the effects of subjective aspects to better reproduce the choice process. Their applications include but are not limited to the analysis of travel mode choice [139–142], and the processes of modelling the choices of route, departure time, vehicle, fuel, and so on [59, 143–146], but also to analyse the effects of soft measures in nudging individuals towards more sustainable means of transportation [147–149].

In general ML algorithms offer a higher flexibility compared to DCMs, since they require less restrictive pre-assumptions on the relations between different variables, which are considered as explanatory for a given phenomenon. At the same time, ML methods are usually black-box methods, which do not allow to extrapolate much information from their outputs, which is instead one of the aims of choice modelling in transportation [27]. To go around this issue, there have already been some attempts in the literature to estimate values for the elasticities and marginal effects connected to a specific variable, to compare them to the results obtained by using similar econometric models, and to have a way of interpreting the outputs of a machine learning model. Some existing studies have already proven the reliability of such procedures [30, 95].

We wanted to also include, among the inputs of our models, psycho-attitudinal indicators, to further differentiate the individuals based on their attitudes, intention, perceptions, and so on. Since ICLV models can be very complex, especially when the number of latent variables considered increases, the specification of each component of the model can be tedious and potentially lead to mistakes from the modeller. ML models could represent a valid alternative, as they do not require researchers to define every detail, since they are data-driven methods and, by definition, they “learn” from the data itself how the phenomenon has developed.

Although there have already been some attempts at producing machine learning models which considered latent variables [35, 123, 150], these studies are characterized by some limits, in the sense that they estimate latent factors on the basis of the socio-economic attributes of the individuals, without taking into account any additional information coming from psycho-attitudinal indicators, whose value originates from specifically engineered questionnaires.

Given the above discussion, in the current chapter we propose an alternative method employing machine learning algorithms for the development of discrete choice models with latent variables to reproduce the phenomenon of mode choice in transportation. Unlike previous models relying on machine learning algorithms, this new model would allow us to use latent variables connected to a series of attitudinal indicators. We thus built anew the same sub-modules which constitute the ICLV model, namely a module to estimate the values of the latent variables, one to integrate the information of the attitudinal indicators through a series of measurement equations, and finally the proper choice model. To build the measurement equations module, we used the already existing Consistent Rank Logits (CORAL) implementation developed by Cao *et al.* [151] and publicly available.

We used this framework to build three different models with different combination of latent variables, and we compared the results to an equivalent version of a more classical ICLV model. We tried to validate our results by using a database containing real data obtained from a revealed preference travel survey conducted among a sample of workers and university students in the metropolitan area of Cagliari in 2019. The dependent variable of our model is the choice to commute by using either car, public transport, or walking. The dataset contains level of service indicators for each alternative, personal and socio-economic characteristics of each individual, and lastly psycho-sociological variables competing to individuals’ attitudes, intentions and perceptions towards the use of sustainable transport modes.

We used 80% of the data to train the models, and the remaining 20% as a testing set and to estimate the final values for each indicator. Since neural networks results are dependent on both the values of their hyper-parameters and on their initialization, we estimated several of them by considering different sets of hyper-parameters to find the optimal values, and then we used those values to verify the stability of the results with different random initializations.

Regarding the outputs, we focused on two aspects: the first, direct and cross-elasticities relative to the level of service variables of the different modal alternatives, to verify the neural network is working as expected, according to the microeconomic theory; the second, a series of pseudo-elasticities [152] connected to the latent variables through the socio-economic attributes used to define them, to measure their effects on the choice probabilities for the different alternatives and compare the results, in terms of impact of psychological variables on choice probabilities, of classical ICLV model and the proposed model.

4.2 Methodology

The level of utility that a decision maker *n* receives from choosing an alternative *j* can be defined as U_{nj}. However, researchers cannot observe this utility directly, but rather they observe one or more aspects $\boldsymbol{x}_{nj}$ which are supposed to influence the utility. Hence, we can only obtain a value for a representative utility $V_{nj} = f(\boldsymbol{x}_{nj})$, and there exists a relation between the two in the form of $U_{nj} = V_{nj} + \varepsilon_{nj}$, where ε_{nj} is a random error component which captures the effects of all the unobserved elements influencing U_{nj}.

The simplest econometric model used to represent choice behaviour is the multinomial logit (MNL) model, which is based on the independence of irrelevant alternatives (IIA), meaning that the relative probability of choosing one alternative over another is independent from the other alternatives [2, 50]. A MNL model is obtained by assuming that the random error components ε_{nj} are independently, identically distributed extreme values. The utility perceived by individual *n* for alternative *j* is a linear combination of observed variables $\boldsymbol{x}_{nj}$ which are combined by using the parameters $\boldsymbol{\beta}$ to represent the tastes of the individuals:

$$V_{nj} = \boldsymbol{\beta}\boldsymbol{x}_{nj} \tag{10}$$

The choice probability for alternative *j* and individual *n* takes then the form:

$$P_{nj} = \frac{e^{V_{nj}}}{\sum_j e^{V_{nj}}} \tag{11}$$

MNL models, like many other even more complex models, are based on microeconomic theory, which tended to consider decision makers as rational self-interested actors engaged in a process of evaluation of both costs and benefits associated with a particular choice, trying to maximize their personal well-being (*i.e.*, utility) given a set of constraints determined by the market. Traditionally, discrete choice models have focused on observable variables, such as product attributes and socioeconomic characteristics of the individuals, and treated consumers as optimizing black boxes with predetermined wants and needs is at odds [153]. However, a deeper understanding of the determinants characterizing mode choice is essential to successfully design environmentally sustainable transport solutions in line with the preferences of different people [154]. Modal choice can be seen as the decision process that leads an individual to choose a specific travel alternative among the different ones available to them. This process can happen both consciously and unconsciously, and its analysis needs to consider a wide range of tools and factors from different disciplines (economy, sociology, geography and psychology) [155]. Socio-economic indicators (*e.g.*, age, gender, education level, occupation, income, household composition, car ownership), spatial characteristics of the environment (*e.g.*, density, diversity, proximity to infrastructures and services, parking availability), and trip attributes (*e.g.*, motivation, distance, travel time, travel cost, departure time, number of stops) have seen a widespread use in the history of choice modelling. However, an approach which considers only these variables, is in direct contrast with the findings in the field of social sciences, which have demonstrated how choice behaviour is also influenced by psychological factors [7]. Thus, in more recent approaches, the inclusion of socio-psychological and attitudinal indicators has allowed researchers specify a new kind of model, which can take into consideration more personal factors, such as experience, lifestyle choices, habits and perceptions [155]. These models of disaggregate decision-making, which include latent constructs for capturing the impact of subjective factors on choice processes, are called hybrid choice models (HCM) or integrated choice and latent variable (ICLV) models. They were originally proposed by McFadden [8] and Train *et al.* [9], but they only became popular much later following the work of Ben-Akiva *et al.* [10], with an increasing number of researchers adopting these models in transportation and travel mode choice [139–142].

4.2.1 The ICLV model

The general formulation of the ICLV needs two main components: a latent variable model and a discrete choice model, as represented in Figure 5. The latent variable model is composed by a set of structural equations, which describe the latent variables (*e.g.,* attitudes, perceptions) in terms of observable characteristics of the individuals, and a group of measurement equations that link the latent variable to the observed indicators (obtained from responses to questions of a survey). By simultaneously integrating both the discrete choice model and the latent variable model, the latent variables can be interpreted as additional explanatory variables included in the utilities of choice alternatives.

The latent variable model requires the specification of the distribution of the latent variable given the explanatory variables. For example, for any individual q:

$$x_q^* = h(\boldsymbol{x}_q; \boldsymbol{\lambda}) + \omega_q, \qquad \omega_q \sim D(0, \Sigma_\omega) \tag{12}$$

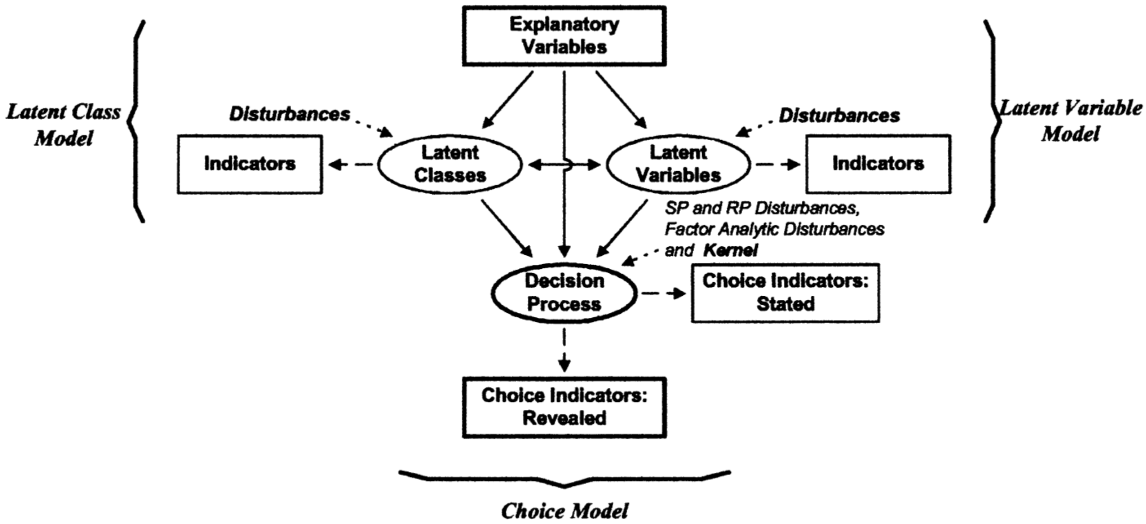


Figure 5. Framework for the Integrated Choice Latent Variable model by Ben-Akiva et al. [10]

where $\boldsymbol{x}_q$ is a vector of explanatory variables, $\boldsymbol{\lambda}$ is a vector of the parameters to be estimated and ω_q is the random error term distributed with variance Σ_ω.

The latent variable model requires the specification of the distribution of the latent variable given the explanatory variables. For example, for any individual q:

$$x_q^* = h(\boldsymbol{x}_q; \boldsymbol{\lambda}) + \omega_q, \qquad \omega_q \sim D(0, \Sigma_\omega) \tag{13}$$

where $\boldsymbol{x}_q$ is a vector of explanatory variables, $\boldsymbol{\lambda}$ is a vector of the parameters to be estimated and ω_q is the random error term distributed with variance Σ_ω. The most common specification for the function h is linear:

$$h(\boldsymbol{x}; \boldsymbol{\lambda}) = \lambda_0 + \sum_{k=1}^{K} \lambda_k x_k \tag{14}$$

For the discrete choice sub-model, the utilities can be defined as:

$$U_q = V(\boldsymbol{x}_q, \boldsymbol{x}_q^*; \boldsymbol{\beta}) + \varepsilon_q, \qquad \varepsilon_q \sim D(0, \Sigma_\varepsilon) \tag{15}$$

where $\boldsymbol{\beta}$ is a vector of the parameters to be estimated and ε_q is the error term distributed with variance Σ_ε. The latent variable model also requires the distribution of the indicators' conditional on the values of the latent variables:

$$I_q = m(\boldsymbol{x}_q, \boldsymbol{x}_q^*; \boldsymbol{\alpha}) + \upsilon_q, \qquad \upsilon_q \sim D(0, \Sigma_\upsilon) \tag{16}$$

where I_q is the reported value, $,\boldsymbol{x}_q^*$ is the vector containing the latent variables, $\boldsymbol{x}_q$ is a vector of explanatory variables, $\boldsymbol{\alpha}$ is a vector of parameters and υ_q is the random error distributed with variance Σ_υ. The measurement equation of the discrete choice model is defined by a dummy variable y_i that assumes the value one if the chosen alternative has the highest utility among all the available alternatives, and zero otherwise:

$$y_i = \begin{cases} 1 & if\ U_i = max_j\{U_j\} \\ 0 & otherwise \end{cases} \tag{17}$$

If no latent variables were considered, the probability function for the choice of alternative *i* would be:

$$P_i(\boldsymbol{y}|\boldsymbol{x}; \boldsymbol{\beta}, \boldsymbol{\Sigma}_\varepsilon) \tag{18}$$

The choice model can take several different forms, *e.g.*, mixed logit, nested logit, probit, ordered probit, ordered logit. In a modelling framework which includes latent variables, if the error components ε and ω are independent, the probability function is obtained by integrating the choice model over the distribution of the latent constructs:

$$P_i(\boldsymbol{y}|\boldsymbol{x}; \boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\Sigma}_\varepsilon, \boldsymbol{\Sigma}_\omega) = \int_{\boldsymbol{x}^*} P_i(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{x}^*; \boldsymbol{\beta}, \boldsymbol{\Sigma}_\varepsilon)\, g(\boldsymbol{x}^*|\boldsymbol{x}; \boldsymbol{\lambda}, \boldsymbol{\Sigma}_\omega) d\boldsymbol{x}^* \tag{19}$$

which is an integral in *M* dimensions, where *M* is the number of latent variables in $\boldsymbol{x}^*$, while g is the density function of the latent variable.

Then, the indicators can be introduced to improve the accuracy of the estimates of the structural parameters. Assuming that all the error components (ε, ω, υ) are independent, the joint probability of observing variables $\boldsymbol{y}$ and $\boldsymbol{I}$ is:

$$P_i(\boldsymbol{y}, \boldsymbol{I}|\boldsymbol{x}; \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\Sigma}_\varepsilon, \boldsymbol{\Sigma}_\upsilon, \boldsymbol{\Sigma}_\omega) = \int_{\boldsymbol{x}^*} P_i(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{x}^*; \boldsymbol{\beta}, \boldsymbol{\Sigma}_\varepsilon) f(\boldsymbol{I}|\boldsymbol{x}, \boldsymbol{x}^*; \boldsymbol{\alpha}, \boldsymbol{\Sigma}_\upsilon)\, g(\boldsymbol{x}^*|\boldsymbol{x}; \boldsymbol{\lambda}, \boldsymbol{\Sigma}_\omega) d\boldsymbol{x}^* \tag{20}$$

The first term of the integrand corresponds to the choice model, the second term and the third term to the measurement equation and structural equation from the latent variable model respectively. Finally, to estimate the unknown parameters of the ICLV model, since the full integrand function is usually very complex, exact integration is not possible, and numerical integration needs to be implemented. The simulated log likelihood, which will be maximized, takes the form:

$$\sum_{q=1}^{Q}\sum_{i=1}^{N}\delta_{qi} ln[P_{qi}(\boldsymbol{y},\boldsymbol{I}|\boldsymbol{x};\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\lambda},\boldsymbol{\Sigma}_{\varepsilon},\boldsymbol{\Sigma}_{\upsilon},\boldsymbol{\Sigma}_{\omega})] \tag{21}$$

where Q is the total number of individuals, N is the set of alternatives, and δ_{qi} is a dummy factor which is equal to 1 if individual q chooses alternative i, and 0 otherwise.

4.2.2 Using neural networks to reproduce the ICLV framework

The new model we are proposing is based instead on a Feed-Forward Neural Network [12] which tries to reproduce the same structure of the ICLV model. Like the hybrid choice model, the structure of the complete neural network can be split in different modules which correspond to a particular function. The general framework of the NN model we wanted to test is shown in Figure 6.

The first module is the one whose purpose is modelling the latent variables (leftmost side in Figure 6). This module takes as input a set of socioeconomic variables $\boldsymbol{Z}_{Lk}$, which might differ based on the latent variable considered, and a set of independent random error components sampled from a given distribution, ω_k, for each of the latent variables we want to build. This module than proceeds to use a neural network for each desired latent variable to generate a linear combination of its inputs through a series of hidden layers, and finally outputs a vector containing the estimated latent variables. This vector will be then used as one of the inputs of the following modules.

The choice model module (upper right in Figure 6) is basically an analogue of the neural network choice model we used to estimate the value of travel time (see Paragraph 3.2). The main difference is that, in this case, among the inputs of the model, other than trip characteristics ($\boldsymbol{X}_k$) and socioeconomic individual attributes ($\boldsymbol{Z}_k$), we have the latent variables outputted by the previous model ($\boldsymbol{LV}$). These inputs go through a series of linear combination neural networks (one for each choice alternative) to produce the alternative specific utility. Finally, the choice probabilities are calculated by using a softmax activation function in the last layer, which leads to the simulation of a negative log likelihood loss. To be similar to the ICLV, the latent variables were introduced in the formulation of the utilities of *K-1* alternatives instead of all of them.

Finally, the ordered logit module (lower right in Figure 6) takes only the latent variables ($\boldsymbol{LV}$) as its inputs, and, after a linear combination, uses them as the input of a particular layer structure (CORAL – consistent rank logits), which was developed by Cao *et al.* [151] to apply ordinal regression using deep neural networks. First of all, given a dataset of ordered observations included in a range of *1* to *K* (*e.g.*, Likert scale responses like in our case), this layer extends the labels into $K-1$ binary labels

$y_i^{(1)}, \dots, y_i^{(K-1)}$, such that the generic label $y_i^{(k)} \in \{0,1\}$ indicates whether y_i is higher than k, *i.e.*, $y_i^{(k)} = 1\ if\ y_i > k$. Using these extended binary labels, the layer trains a single NN with binary classifiers in the output layer. The outputs of the CORAL layer ($\theta_1, \dots, \theta_K$) are then used to estimate the CORAL loss function:

$$-\sum_{q=1}^{Q}\sum_{k=1}^{K-1}\lambda^{(k)}[ln\,(\sigma(\theta_k)y_q^{(k)} + ln\,(1-\sigma(\theta_k))(1-y_q^{(k)})] \tag{22}$$

where $\sigma(\theta_k)$ represents the logistic sigmoid function:

$$\sigma(\theta_k) = \frac{1}{1+e^{-\theta_k}} \tag{23}$$

and $\lambda^{(k)} > 0$ denotes the weight of the loss associated with the *k*-th classifier, which could be used in the case when some of the classifiers may be less robust or harder to optimize (we assumed $\lambda^{(k)} = 1$ in all cases). Finally, all the losses (one for the choice model module, and *K* from the ordered model module) were combined in a single global loss function, which had to be minimized through back-propagation. To emulate the integration happening in the ICLV model, we calculated the inputs of the loss functions for every draw $\boldsymbol{\omega}^{(r)}$ from the distribution of random parameters, and we obtained the Montecarlo integral approximation [156] by averaging them.

4.2.3 Disaggregate indicators

Regarding the results of the models, we first analysed, to use them as a benchmark to verify the correct behaviour of the models, some elasticities derived by a variation of the level of service variables associated with each travel alternative, obtained by applying the formula (see also Paragraph 3.2):

$$\frac{1}{N}\sum_{n=1}^{N}\left\{\frac{P(\boldsymbol{x}_n+\Delta Z_n)_j - P(\boldsymbol{x}_n)_j}{\Delta Z_n}\cdot\frac{Z_n}{P(\boldsymbol{x}_n)_j}\right\} \tag{24}$$

where $P(\boldsymbol{x}_n)_j$ is the probability of choosing alternative *j*, $\boldsymbol{x}_n$ is a vector containing all the explanatory variables, $Z_n \in \boldsymbol{x}_n$ is the variable for which we are interested in calculating the elasticity, and *N* is the size of the sample.

We also evaluated the effects of the latent variables on the choice probabilities by computing a set of pseudo-elasticities [152], which are estimated with the simple relation:

$$\frac{1}{N}\sum_{n=1}^{N}\{P(\boldsymbol{x}_n+\Delta Z_n)_j - P(\boldsymbol{x}_n)_j\} \tag{25}$$

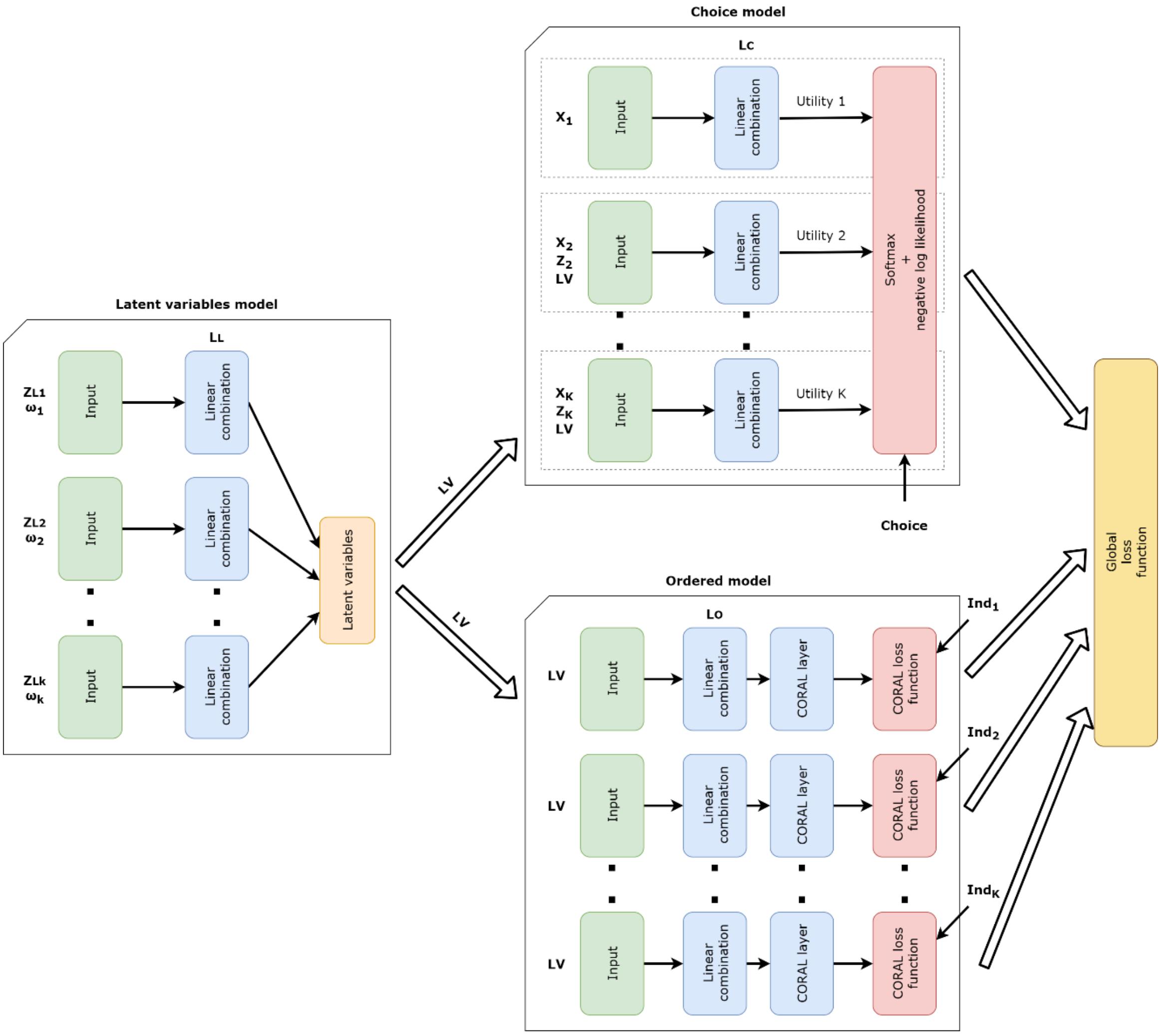


Figure 6. Proposed neural network architecture to emulate an ICLV model.

These represent the absolute (as opposed to the relative one used in the elasticities) variation in the probability when a variable represented by discrete quantities changes its value. In our case, Z_n represents the socio-economic variables we used to build the latent variables. We only evaluated the indirect effects of such variations, by only considering the new value when estimating the value of the latent variable, but not when calculating the utilities associated with the alternatives. All the variables we used when building our latent variables (see Table 19) were dummy variables, with the exception of *Age* and *Nr. of cars*. For the dummy variables we obtained the pseudo-elasticities by estimating the probabilities in the case where each variable assumed value zero for all individuals, while for *Nr. of cars* we increased its value by 1 unit for every household, and *Age* was increased instead so that everyone was hypothesised to be 5 years older.

4.3 Data analysis

The dataset we used (*Svolta Cagliari*) was collected in 2019-2020 during an experimental programme lead by the government of the metropolitan city of Cagliari (Italy), in collaboration with the University of Cagliari. The aim of the program was that of finding possible ways of convincing frequent commuters to change their travel behaviour towards more sustainable means of transport. An online survey was distributed among the population of interest, and it ultimately led to over 4,000 complete questionnaires, although we analysed a sub-sample of the people which travelled by car, PT or walking, and had at least two of these alternative available.

The dataset is composed of 2,873 observations and it is the same used in the chapter in which we compared methods of obtaining the value of travel time (see Paragraph 3.3.2). However, since we were interested in using psycho-attitudinal indicators in our models, we also had to include some additional observations for each individual, pertaining to some questions aimed at investigating individuals' attitudes, intentions and perceptions. Table 17 contains an analysis of these variables, and it shows the item proposed in the online questionnaire (or rather, the closest English translation since the survey was originally in Italian, as shown in Appendix A). Each of the items required a the respondent to choose among a 5-options Likert scale [157] (1 = strongly disagree, 2 = somewhat disagree, 3 = neither agree nor disagree, 4 = somewhat agree, 5 = strongly agree), for several different topics, mostly connected to sustainable means of transport.

The first group of items regarded the individuals' intentions to use sustainable transport modes / not use their car in the following days [158]. The average response on the first item *During the next two weeks I intend to use sustainable transport modes instead of the car (alone)* was 3.36 out of 5, with most people answering *5* (28%) and *4* (24%). Regarding the second item in this group, *During the next two weeks I intend to use a private car*, we observed an average of 2.67 out of 5, and the most popular responses were *1* (27%) and *2* (26%). The last intention-related item, *I am interested in using sustainable transport modes during the next two weeks*, had a relatively higher average response (3.97 out of 5), and almost half (46%) of the respondents chose option *5*. These responses together depict a scenario in which most respondents manifested their intention to use sustainable alternatives, while at the same time not giving up on their cars.

The second group was instead composed of those items aimed at analysing the perceived behavioural control (PBC) concerning sustainable transport modes [159, 160]. The first item was *It would be easy for me to use sustainable transport modes*, and the average response was 3.16 out of 5. The most common responses were *4* (24.5%) and *5* (24.2%). The responses to the second one, *I am certain I can*

Table 17. Analysis of the responses to the psycho-attitudinal survey questions

Questionnaire item	**Name**	**Avg.**	**1**	**2**	**3**	**4**	**5**
During the next two weeks I intend to use sustainable transport modes instead of the car (alone).	Int1	3.36	14.9%	13.9%	19.6%	23.7%	27.9%
During the next two weeks I intend to use a private car.	Int2	2.67	25.7%	27.1%	18.9%	11.5%	16.8%
I am interested in using sustainable transport modes during the next two weeks.	Int3	3.97	5.1%	5.7%	22.5%	20.4%	46.4%
It would be easy for me to use sustainable transport modes.	PBC1	3.16	18.4%	19.7%	13.2%	24.5%	24.2%
I am certain I can use sustainable transport modes during the next week.	PBC2	3.32	19.7%	15.1%	12.5%	18.6%	34.1%
Using sustainable transport modes is possible for me.	PBC3	3.73	9.8%	14.5%	14.0%	16.4%	45.4%
To me, using sustainable transport modes instead of a private car is USEFUL.	Att1	4.16	3.3%	5.1%	10.3%	34.8%	46.5%
To me, using sustainable transport modes instead of a private car is (or would be) PLEASANT.	Att2	3.55	8.1%	12.7%	21.1%	32.2%	25.9%
To me, using sustainable transport modes instead of a private car is RIGHTEOUS.	Att3	4.26	1.3%	2.4%	13.4%	35.2%	47.7%
Most people I know think I should use sustainable transport modes instead of a private car.	NoS1	2.64	23.0%	20.5%	33.2%	16.3%	7.0%
Most people I know use sustainable transport modes instead of a private car.	NoS2	2.33	29.2%	34.4%	17.0%	13.0%	6.3%
I feel a moral obligation to use sustainable transport modes regardless of what everybody else does.	NoM1	3.52	8.7%	9.9%	26.6%	30.2%	24.7%

1 = strongly disagree, 2 = somewhat disagree, 3 = neither agree nor disagree, 4 = somewhat agree, 5 = strongly agree

use sustainable transport modes during the next week, produced an average of 3.32 out of 5, and more than a third of the sample (34%) answered *5*. The third PBC item was *Using sustainable transport modes is possible for me*, and the average response was 3.73 out of 5, while the most chosen option was still *5* (45%). Overall, this groups shows how the respondents perceived sustainable alternative as a possible option for their trip, but this option is perceived as always feasible.

Another group included those items which analysed the respondents' personal attitudes towards the use of sustainable transport modes [161]. The first of this group was *To me, using sustainable transport modes instead of a private car is USEFUL*, the average was 4.16 out of 5 and most responses were either *5* (46%) or *4* (35%). The second attitudinal item, *To me, using sustainable transport modes instead of a private car is (or would be) PLEASANT*, gave an average response of 3.55 out of 5. The most chosen option was *4* (32%), followed by *5* (26%). The last item of this group was *To me, using* *sustainable transport modes instead of a private car is RIGHTEOUS*, the average response was 4.26 out of 5, with the most chosen options being *5* (48%) and *4* (35%). This group expresses the fact that, although most people feel that using sustainable alternatives would be the most correct choice, it would not also necessarily be the most desirable. The final group of items included all those questions pertaining social norms and moral norms linked to the use of sustainable transport alternatives [162].

The first item was *Most people I know think I should use sustainable transport modes instead of a private car*. The average response was 2.64 out of 5, and most respondents chose options *3* (33%) and *1* (23%). The second normative question, *Most people I know use sustainable transport modes instead of a private car*, produced an average of 2.33 out of 5, with the most frequent answer being *2* (34%) and *1* (29%). The third item, which was also the only moral norm one, was *I feel a moral obligation to use sustainable transport modes regardless of what everybody else does*, and its average was 3.52 out of 5. The most chosen option was *4* (30%), closely followed by *3* (27%) and *5* (25%). These items might be interpreted with the fact that most people do not feel highly pressured by their peers to use sustainable transport modes, but rather their choice would be guided by their own belief that these alternatives are the most moral choice.

To identify the latent variables we wanted to build by using these psycho-attitudinal items collected from the survey, we used factor analysis [163]. To measure sample adequacy of the different constructs, we used the Kaiser-Meyer-Olkin test (KMO) [164]. Since the questions were already based on underlying behavioural theories exploratory factor analysis was unnecessary, so we directly performed a confirmatory factor analysis, by keeping together the items concerning the same (or similar) behavioural aspects, *i.e.*, intentions, PBC, attitudes, and norms. The first latent variable (LV1) is thus *Intentions to use sustainable transport modes* (KMO = 0.652), the second one (LV2) *Perceived behavioural control for sustainable transport modes* (KMO = 0.732), the third latent construct is (LV3) *Attitude towards sustainable transport modes* (KMO = 0.674), and the fourth and final one (LV4) *Behavioural norm* (KMO = 0.564). This means that the KMO for LV4 is below the threshold value (0.6) for it to be considered reliable. The main results of the factor analysis are shown in Table 18, where we show the factor loadings for each survey item and the Cronbach's alpha [165] for each latent variable. The value of the alpha is acceptable (≥ 0.7 [166]) for all the latent factors, except for, once again, LV4, which will be thus excluded. Ultimately, we also chose to consider just two latent variables in the models, and we opted for LV1 and LV2 based on values of the Cronbach's alphas.

Table 18. Confirmatory factor analysis for the psycho-attitudinal indicators

Latent factor	Variables	Loadings	Cronbach's alpha
LV1 – *Intentions to use sustainable transport modes*	Int1	0.898	0.7781
	Int2	0.756	
	Int3	0.561	
LV2 – *Perceived behavioural control for sustainable transport modes*	PBC1	0.878	0.8914
	PBC2	0.913	
	PBC3	0.778	
LV3 – *Attitude towards sustainable transport modes*	Att1	0.691	0.7058
	Att2	0.612	
	Att3	0.741	
LV4 – *Behavioural norm*	NoS1	0.764	0.5005
	NoS2	0.473	
	NoM1	0.316	

4.4 Model specification

Like we did in the case of the value of travel time estimation, we had to define which variables to use in the models (see Paragraph 3.4.1). In this case, we also decided to estimate an ICLV model beforehand to verify that the variables we chose did not generate any issues during the estimation. Then we used the same variables when defining the inputs of the neural networks, so that the econometric information we were going to extract would have been comparable directly. Since we had two latent variables to use, we also decided to produce different models based on the variable considered. We thus built three different models, one each for LV1 and LV2, and another one in which both latent variables were considered. The parameters associated with the variables, and the corresponding t-stats we obtained by estimating the three models are shown in Table 19. All the results were obtained by using PythonBiogeme [132].

Once again, since neural networks are generally non-identifiable (see Paragraph 3.2), we tried to mitigate the risks of obtaining skewed results by identifying the ideal set of hyperparameters to use in our final models. To do so, we trained several neural networks, modifying one hyperparameter at a time and analysing the effects this change had on the results. In particular, this time we focused on the number of epochs the model would run for, the number of hidden nodes in the NNs composing the choice model, and the number of hidden nodes in the linear combination layer preceding the CORAL layer. We chose to keep the number of hidden layers as low as possible, *i.e.*, one in the choice model, two in the ordered models (one for linear combination and one for CORAL), to keep the model as simple as possible and because we also saw in our previous experiments (Paragraph 3.4.2) that additional layers usually lead to worse results. For each hyperparameter, we then chose, among those we tested, the lowest value for which the results seemed to stabilise around an average value.

Table 19. Modelling results obtained with PythonBiogeme for the ICLV models

		Model 1 (LV1)		Model 2 (LV2)		Model 3 (LV1+LV2)	
	Variable name	**Value**	**Rob. t-test**	**Value**	**Rob. t-test**	**Value**	**Rob. t-test**
Car attributes	Cost	-0.222	-1.93	-0.223	-1.70	-0.236	-1.88
	Time	-0.033	-1.12	-0.052	-1.74	-0.036	-1.18
PT attributes	ASC	-0.297	-0.39	-1.240	-1.56	-2.070	-2.56
	Cost	-0.160	-1.08	-0.265	-1.65	-0.153	-0.99
	Time	-0.019	-2.54	-0.029	-3.91	-0.024	-3.25
	Age	-0.155	-1.50	-0.319	-3.24	-0.190	-1.82
	Gender = male	-0.450	-2.51	0.006	0.03	-0.145	-0.80
	Presence of children	-0.087	-0.74	-0.178	-1.39	-0.040	-0.31
	Owns a car	-2.720	-5.13	-1.670	-3.21	-1.900	-3.49
	Student	1.430	5.15	1.490	5.41	1.420	5.04
	House in Cagliari city	-0.525	-2.35	-0.081	-0.36	-0.581	-2.53
	Nr. of cars	-0.357	-3.12	-0.411	-3.44	-0.315	-2.67
	LV1 (PBC)	0.836	10.54	-	-	0.528	8.18
	LV2 (Intention)	-	-	0.908	9.56	0.547	7.62
Walking attributes	ASC	-1.850	-1.13	-2.450	-1.36	-3.430	-2.06
	Time	-0.136	-9.74	-0.140	-10.13	-0.137	-9.74
	Age	0.256	1.92	0.064	0.48	0.225	1.58
	Gender = male	-0.237	-0.96	0.193	0.81	-0.014	-0.06
	Presence of children	0.189	1.14	0.127	0.74	0.227	1.28
	Owns a car	-2.410	-4.22	-1.450	-2.57	-1.720	-2.89
	Student	0.796	2.05	0.896	2.30	0.819	2.04
	House in Cagliari city	1.080	0.92	1.890	1.43	1.120	1.03
	Nr. of cars	-0.177	-1.09	-0.238	-1.36	-0.136	-0.80
	LV1 (PBC)	0.864	8.62	-	-	0.608	7.05
	LV2 (Intention)	-	-	0.836	7.66	0.447	5.07
Ordered models	Delta1 PBC1	2.640	18.33	-	-	2.700	17.47
	Delta2 PBC1	1.560	15.79	-	-	1.590	15.28
	ASC PBC2	-0.440	-2.59	-	-	-0.372	-2.33
	Beta PBC2	1.350	9.65	-	-	1.240	9.40
	Delta1 PBC2	2.570	12.63	-	-	2.440	12.94
	Delta2 PBC2	2.000	12.08	-	-	1.880	12.30
	ASC PBC3	1.540	14.60	-	-	1.550	14.70
	Beta PBC3	0.670	15.03	-	-	0.653	14.05
	Delta1 PBC3	1.910	16.91	-	-	1.900	16.84
	Delta2 PBC3	1.350	17.53	-	-	1.350	17.43
	Delta1 Int1	-	-	1.990	13.17	1.880	12.61
	Delta2 Int1	-	-	2.160	15.02	2.030	13.89
	ASC Int2	-	-	-1.420	-8.92	-1.550	-7.47
	Beta Int2	-	-	0.895	9.37	1.000	7.46
	Delta1 Int2	-	-	2.680	19.46	2.760	16.01
	Delta2 Int2	-	-	2.020	17.57	2.080	14.91
	ASC Int3	-	-	2.040	19.91	2.010	19.42
	Beta Int3	-	-	0.364	12.13	0.393	11.83
	Delta1 Int3	-	-	0.939	12.11	0.939	12.11
	Delta2 Int3	-	-	1.750	22.81	1.750	22.66
LV1 (PBC)	ASC	7.340	11.65	-	-	7.430	11.31
	Age	-0.333	-3.58	-	-	-0.334	-3.52
	Gender = male	0.037	0.21	-	-	0.092	0.48
	Graduate	-0.408	-2.12	-	-	-0.417	-2.00
	Presence of children	-0.489	-4.31	-	-	-0.500	-4.13
	Owns a car	-2.100	-6.73	-	-	-2.280	-6.28
	Student	1.330	4.44	-	-	1.280	4.15
	House in Cagliari city	2.150	10.00	-	-	2.150	9.41
	Nr. of cars	-0.784	-6.43	-	-	-0.711	-5.74
	Dep. Time 8:30-9:29 AM	-0.367	-2.06	-	-	-0.356	-1.84
	Sigma	3.610	18.26	-	-	3.710	17.11
LV2 (Intention)	ASC	-	-	7.540	12.440	7.080	11.57
	Gender = male	-	-	-0.566	-3.510	-0.539	-3.54
	Presence of children	-	-	-0.451	-4.640	-0.385	-4.08
	Owns a car	-	-	-3.370	-9.250	-3.200	-8.76
	Student	-	-	1.590	7.190	1.530	7.05
	House in Cagliari city	-	-	1.300	6.900	1.150	6.40
	Nr. of cars	-	-	-0.654	-5.530	-0.588	-5.12
	Sigma	-	-	3.260	13.040	3.030	11.72

Table 20 shows the definitive values considered during the estimation of the final results. For additional information on the results that we analysed to get to these values, see Appendix C. Keep in mind that we only performed the analysis only using one model (the one which considers only LV1), and we assumed the same results could be extended to the other models. The neural network models were estimated using the PyTorch library for Python [138].

Table 20. Hyperparameters sets for all models

N. epochs	N. hidden nodes (choice model)	N. hidden nodes (latent model)
1000	200	20

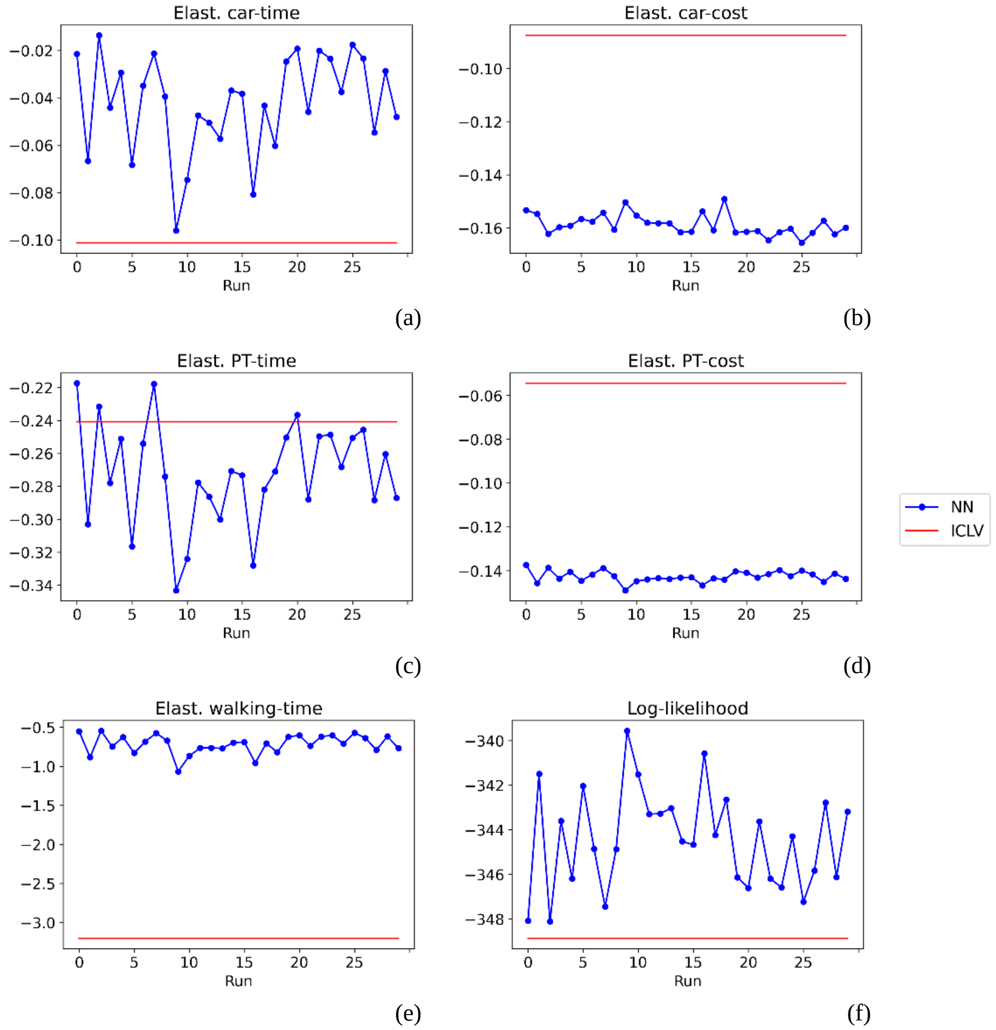
4.5 Results

In the following pages we are going to show the results we obtained by using our neural network architecture and compare them to the ones we got with Biogeme for the ICLV. We trained a total of 30 NNs for each model specification by using 80% of the dataset. We then observed the trend of some benchmark indicators (*i.e.*, direct elasticities for the level of service variables and log-likelihood) estimated using a 20% testing set to check if they were stable and compared them with the values obtained by using the corresponding ICLV. As we said previously, we first analysed, as a benchmark, the elasticities with respect to the level of service variables, and then the pseudo-elasticities associated with the latent variables.

4.5.1 Model 1 - LV1 (PBC)

Figure 7 shows the benchmark results for the neural network trying to reproduce the ICLV with only one latent variable (LV1 - PBC). As expected, most graphs show a regular trend, even if the values are quite different from those that we obtained with the ICLV. The most regular values are those of the elasticities for alternative *car* w.r.t. travel cost, for alternative *PT* w.r.t. travel cost, and for *walking* w.r.t. travel time (even if the values for the latter are probably the furthest ones from those of the ICLV). Elasticities w.r.t travel time for both *car* and *PT* present a wider range of values, but they still can be considered regular, and their values do not show extreme differences compared to the ICLV results. The log-likelihood was higher than the one we obtained for the ICLV for all NNs, so we achieved an overall slightly better level of fit.

Table 21 shows instead all the direct- and cross-elasticities for all modes and all alternatives for both our neural network (NN) and the ICLV, generated by a positive variation in the level of services variables. The values for the NN were obtained by averaging those we got from the 30 different models we trained.



(a) elasticity of car choice probability w.r.t. car travel time; (b) elasticity of car choice probability w.r.t. car cost; (c) elasticity of PT choice probability w.r.t. PT travel time; (d) elasticity of PT choice probability w.r.t. PT cost; (e) elasticity of walking choice probability w.r.t. walking travel time; (f) log-likelihood

Figure 7. Neural network benchmark results for Model 1 (LV1)

Table 21. Elasticities for level of service variables with Model 1 (LV1)

Variable	Model	Car	PT	Walking
Car cost	**NN**	-0.159	0.230	0.023
	ICLV	-0.088	0.134	0.035
Car travel time	**NN**	-0.042	0.064	0.012
	ICLV	-0.101	0.160	0.076
PT cost	**NN**	0.082	-0.143	0.036
	ICLV	0.034	-0.055	0.073
PT travel time	**NN**	0.148	-0.272	0.056
	ICLV	0.151	-0.241	0.286
Walking travel time	**NN**	0.089	0.126	-0.717
	ICLV	0.226	0.689	-3.205

We should first notice how all the signs we obtained are consistent with the microeconomic theory, *i.e.*, all direct elasticities are negative, while all cross-elasticities are positive, so we found no problems at least in the regard. We will now analyse them by grouping them based on the corresponding variable affecting the choice probability:

- *Car cost*: the direct elasticity for the NN (-0.16) is roughly double of that given by the ICLV (-0.09), and a similar pattern can be seen in the cross-elasticities for the *PT* alternative (0.23 for the NN and 0.13 for ICLV), while the cross-elasticity for the *walking* alternative is lower for the NN (0.02 *vs.* 0.035); both models however are giving an overall similar information, that is, the demand show very low elasticity towards variation of the travel costs of the alternative *car*;
- *Car travel time*: in this case, instead, the values estimated with the NN are much lower than those obtained with the ICLV, since the direct elasticities are -0.04 (NN) and -0.10 (ICLV), while the cross elasticity for *PT* is 0.06 for the NN and 0.16 for the ICLV, and the ones for *walking* are 0.01 (NN) and 0.08 (ICLV); it is thus clear that, also in this case, the two modelling framework agree in the fact that the probability of choosing any of the three alternatives is relatively inelastic towards variations of *car travel time*;
- *PT cost*: like in the case of *car cost*, we get values twice as big for the NN compared to the ICLV, with cross elasticities for *car* being 0.08 (NN) and 0.03 (ICLV), although in this case the direct elasticity is almost triple for the NN (-0.14 vs. 0.06); also for this variable, NN results are lower for *walking*, since the cross-elasticity is 0.04 for the NN and 0.07 for the ICLV model; once again, despite some differences, both models describe a scenario in which the demand behaves inelastically when introducing changes in the cost sustained to use PT;
- *PT travel time*: contrary to previous cases, NN and ICLV gave us very similar results for direct elasticities (-0.27 and -0.24 respectively) and cross-elasticities for alternative car (0.15 in both cases), however the alternative *walking* is still an exception, since the value we obtained with the NN is 0.06, and the one of the ICLV is much higher (0.29); nonetheless, both models are still agreeing in the fact that the choice probabilities are inelastic towards variations of the travel time for the *PT* alternative
- *Walking travel time*: finally, these last elasticities present probably the most striking differences among all these results; as a matter of fact, while the values for the *car* alternative cross-elasticity are relatively low, 0.09 for the NN and 0.23 for the ICLV, we can already see a problem for *PT*, whose cross-elasticities are 0.13 (NN) and 0.69 (ICLV) and describe thus a different phenomenon, with the ICLV coming closer to an elastic behaviour; but the direct

elasticities are even more discordant, since the NN gave us a value of -0.72 (still lower than 1) and the ICLV -3.20 instead, which would mean the NN is severely underestimating the effects of a variation in the travel time when compared to the results given by the ICLV.

Table 22 finally shows the pseudo-elasticities, describing the effects of variations of the latent variable LV1 (PBC) on the choice probabilities of the three alternative means of transport. We again compare the results of the ICLV model to those of the neural network, whose values were obtained again by averaging those we got from the 30 different models we trained. In this case, there are no microeconomic theories which can tell us exactly if a positive or negative sign is correct or not beforehand, so we cannot comment on that aspect. We will again analyse the pseudo-elasticities by grouping them based on the corresponding modified variable.

- *Age*: for an increase of 5 years in the age of the sample, regarding the *car* alternative, both models predict very similar outcomes, that is an increase between 1.0% (NN) and 1.2% (ICLV) in the probabilities; a similar result is shown for *PT*, even though the values are more spread out, with -1.4% for NN and -0.9% for ICLV; *walking* performs inconsistently instead, since the NN predicted a positive value (0.3%) and the ICLV a negative one (-0.4%), however, since they are both relatively low values, we could assume this happens because they are fluctuating around zero;

Table 22. Pseudo-elasticities for LV1 with Model 1 (LV1)

Variable	Model	Car	PT	Walking
Age	**NN**	1.00%	-1.37%	0.35%
	ICLV	1.20%	-0.90%	-0.40%
Gender = male	**NN**	0.57%	-0.71%	0.20%
	ICLV	0.10%	-0.10%	0.10%
Graduate	**NN**	-1.60%	2.09%	-0.51%
	ICLV	-1.70%	1.20%	0.60%
Presence of children	**NN**	-1.13%	1.43%	-0.30%
	ICLV	-1.50%	1.00%	0.70%
Owns a car	**NN**	-15.80%	18.74%	-2.94%
	ICLV	-15.80%	11.70%	5.10%
Student	**NN**	3.65%	-5.36%	1.72%
	ICLV	3.50%	-2.90%	-0.60%
Dep. Time 8:30-9:29 AM	**NN**	-1.10%	1.46%	-0.33%
	ICLV	-1.20%	0.80%	0.40%
House in Cagliari city	**NN**	5.51%	-10.26%	4.75%
	ICLV	9.20%	-6.00%	-4.80%
Nr. of cars	**NN**	4.61%	-6.37%	1.76%
	ICLV	5.70%	-4.20%	-1.90%

- *Gender = male*: in the case in which all the sample is composed by females, both models predict very small difference compared to the actual situation, since for *car* we have 0.6% (NN) and 0.1% (ICLV), for *PT* -0.7% (NN) and -0.1%(ICLV), and for *walking* 0.2% (NN) and 0.1% (ICLV);
- *Graduate*: if everyone in the sample lacked an higher education level, we would see some shifts in the modal split, since both models predict a decrease in the probability of choosing the car (-1.6% for NN and -1.7% for ICLV) and an increase in the use of public transport (2.1% for NN and 1.2% for ICLV); however, the models disagree again in their predictions for the alternative *walking*, since the NN says there would be a decrease (-0.5%) and the ICLV foresees an increase (0.6%), but we could again reconduct this discrepancy to the fact they are both very close to zero;
- *Presence of children*: should none of the households include children among their members, the models predict a decrease in the probabilities of choosing the *car* alternative (-1.1% and -1.5% for NN and ICLV respectively) and an increase in those of using *PT* (1.4% and 1.0%); once again we got contradicting results for *walking* (-0.3% for the NN and 0.7% for the ICLV), so the same comments we made for previous variables are still valid;
- *Owns a car*: if no one among the respondents in the testing set owned a car, according to the models we would see significant changes in the modal split; as one would probably expect, the probability of choosing *car* as a means of transport sharply declines (-15.8% for both models), while at the same time *PT* becomes much more popular (18.7% for the NN, 11.7% for the ICLV); however, in this case the discrepancy in the outputs for the *walking* alternative becomes very evident, since the NN predicts a decrease in this option (-2.9%) while the ICLV shows a significant increase (5.1%), and in this case we cannot assume a fluctuation in the values (like we did for previous variables) since they are quite different from zero;
- *Student*: in the case when none of the individuals were students, both NN and ICLV state that the car alternative would see an increase in its choice probability (3.6% and 3.5%), while public transport would decrease by -5.4% for the NN and -2.9% for the ICLV; while less pronounced than in the case of car ownership, the issue of opposing signs for *walking* is still present, with NN predicting a 1.7% increase and ICLV a -0.6% decrease;
- *Dep. Time 8:30-9:29 AM*: if everyone decided to leave their house outside of the peak-hour starting at 8:30 AM, we could observe a slight decrease in the use of cars (-1.1% for NN and -1.2% for ICLV) and an increase of the probability to choose PT as a means of traveling (1.5%

and 0.8%); the outcomes predicted for *walking* are still opposite to each other (-0.3% for NN and 0.4% for ICLV), but we can consider them negligible like in previous cases;

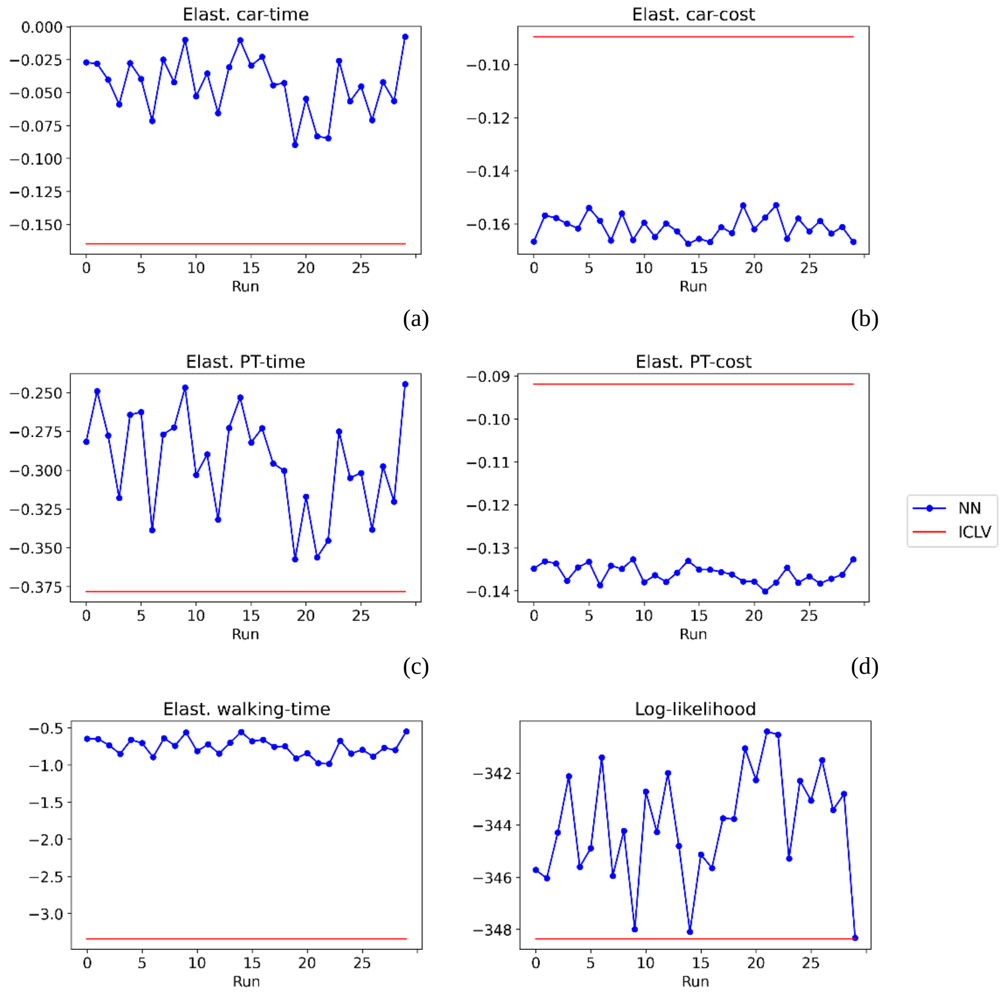
- *House in Cagliari city*: were all the people living outside of the city of Cagliari, and thus further away from their workplace (or university in the case of students), both models predict an increase in the use of cars, quantified in 5.5% for the NN and 9.2% for the ICLV; they also show there would be a significant change in the use of public transport, -10.3% according to the NN and -6.0% for the ICLV; the results for walking are once again contradictory, with the NN predicting an increase of 4.7% and the ICLV a decrease of -4.8%, and since they are not negligible we cannot say with certainty which of the two models is closer to the real outcome;
- *Nr. of cars*: if everyone had one more car at their disposal in their household, both models, quite unsurprisingly, predict an increase in the probability of choosing *car* as a means of transport (4.6% for the NN and 5.7% for the ICLV model), and a decrease in the use of public transport (-6.4% and -4.2%); also in this last case, the predicted outcomes for walking show a discrepancy, since the NN says there would be an increase (1.8%) and ICLV instead shows a decrease (-1.9%), and regretfully we cannot disregard them.

In general, we can conclude that the neural network performs comparably to the ICLV, with the exception of the pseudo-elasticities for the *walking* alternative, which seem to be interpreting the inverse phenomenon. Also, we cannot affirm which of the two models is reproducing more closely the real outcomes generated by changes in the socio-economic variables we use to define the latent variable, and in the worst-case scenario we could also have to consider that both are making wrong predictions. One possible explanation of these discrepancies is that, while both models were theoretically defined following the same logic, it is possible that using different optimization algorithms, we could end up with a different laten variable interpretation in each modelling framework. Thus, the latent variables would be representing different aspects and should be interpreted differently, making it difficult to compare them.

4.5.2 Model 2 – LV2 (Intentions)

Figure 8 shows the benchmark results for the neural network models following the ICLV structure and using the latent variable LV2 (Intentions). Also in this case, most of the graphs show regular patterns, with the predicted values being very close to their average. The same comments we made for Model 1 (Paragraph 4.5.1) can be extended to these graphs, since the elasticities for alternative car and cost, for PT and cost, and for walking and time kept their almost constant trend, while the others (car and time, PT and time) present a wider range of values. The average log-likelihood of the NNs is still slightly higher than the one we got from the ICLV models.

Table 23 shows the direct- and cross-elasticities we obtained by using the Model 2 specification with the NN and the ICLV, when a variation is introduced in the level of services variables. The values we obtained are very close to those we got in the case of Model 1 (Table 14), and they still maintain the same relations, so the comments we previously made are still valid.



(a) (b) (c) (d) (e) (f)

(a) elasticity of car choice probability w.r.t. car travel time; (b) elasticity of car choice probability w.r.t. car cost; (c) elasticity of PT choice probability w.r.t. PT travel time; (d) elasticity of PT choice probability w.r.t. PT cost; (e) elasticity of walking choice probability w.r.t. walking travel time; (f) log-likelihood

Figure 8. Neural network benchmark results for Model 2 (LV2)

Table 23. Elasticities for level of service variables with Model 2 (LV2)

Variable	Model	Car	PT	Walking
Car cost	**NN**	-0.161	0.234	0.024
	ICLV	-0.090	0.137	0.041
Car travel time	**NN**	-0.044	0.066	0.012
	ICLV	-0.165	0.255	0.139
PT cost	**NN**	0.078	-0.136	0.034
	ICLV	0.057	-0.092	0.117
PT travel time	**NN**	0.160	-0.295	0.060
	ICLV	0.233	-0.378	0.426
Walking travel time	**NN**	0.092	0.131	-0.754
	ICLV	0.287	0.723	-3.341

Table 24 shows the values associated with the pseudo-elasticities which predict the effects on the choice probabilities caused by variations in the latent variable LV2 (Intentions), generated in turn by a change in the socio-economic variables we used to define LV2. We compare the results of the ICLV model to those we obtained by averaging the results from 30 different neural network models we trained with the same hyperparameters but different initializations. Again, there are no microeconomic theories to give us any hint to know beforehand if the results we obtained are correct or not, and we will limit ourselves to a comparison, and analyse the pseudo-elasticities considering each socio-economic variable one at a time.

- *Gender = male*: if we hypothesise everyone in our sample was an individual of female gender, the models predict that the probability of using the car alternative would decrease, but while the NN shows a very limited change (-0.3%), the ICLV says instead the variation would be much higher (-2.1%); the prediction for the walking alternative has the two model mostly agreeing (0.6% for NN and 0.7% for ICLV), but public transport generated some issues (not unlike the ones we encountered for model 1 and walking – Paragraph 4.5.1), since the NN predicts a small decrease (-0.3%), while the ICLV says the probability would increase by a relatively larger amount (1.5%);

- *Presence of children*: in the case where no children were present in the households of the respondents, the models behave very similarly to the case of the previous variable; as a matter of fact, for the *car* alternative the NN says there would be a -0.31% decrease, and the ICLV predicts one of -1.5%; for *walking* we have a ~0.5% increase for both models, and once again we encounter a discrepancy for *PT*, since the NN predicts a very small decrease (-0.2%) and the ICLV model an increase of 1.1%;

Table 24. Pseudo-elasticities for LV2 with Model 2 (LV2)

Variable	Model	Car	PT	Walking
Gender = male	**NN**	-0.32%	-0.26%	0.58%
	ICLV	-2.09%	1.54%	0.69%
Presence of children	**NN**	-0.31%	-0.16%	0.47%
	ICLV	-1.49%	1.10%	0.49%
Owns a car	**NN**	-4.99%	-2.73%	7.72%
	ICLV	-27.18%	21.29%	6.57%
Student	**NN**	0.27%	0.76%	-1.04%
	ICLV	4.63%	-4.02%	-0.32%
House in Cagliari city	**NN**	1.09%	1.35%	-2.44%
	ICLV	6.26%	-4.43%	-2.50%
Nr. of cars	**NN**	0.69%	0.85%	-1.53%
	ICLV	5.17%	-4.03%	-1.29%

- *Owns a car*: if no one in our sample owned a car, the two models predict very different scenarios, at least for the *car* alternative, for which the NN predict a decrease in its use of -5.0% while the decrease predicted by the ICLV is much larger (-27.2%), and for the *PT* alternative, whose variations are not only very different in value (the NN result is almost ten times smaller than the one from the ICLV model), but also in sign (-2.7% for NN and 21.3% for ICLV); *walking* is an exception, since the values from the two models are very similar, 7.7% for the NN and 6.6% for the ICLV;
- *Student*: supposing that no one among the individuals was a student, we can observe that the NN predicts an increase of 0.3% and the ICLV model one of 4.6% for the probability of choosing *car* as a mean of transport; the choice of walking would decrease by -1.0% according to the NN and by -0.3% for the ICLV; public transport keeps giving conflicting results, with the NN predicting a small increase (0.8%) and the ICLV model saying the share for this mode would decrease by -4.0%;
- *House in Cagliari city*: if no one lived in the city of Cagliari, but in the surrounding areas, the NN says we would observe an increase of 1.1% in the use of *car* and the ICLV instead one of 6.3%, while *walking* would decrease by similar amounts for both models (-2.4% for the NN

and -2.5% for the ICLV model); once again, the models disagree for the *PT* alternative, since NN predict an increase of 1.4% and the ICLV a decrease of -4.4%;

- *Nr. of cars*: in the case where every household had one more car at their disposal, the NN says we would have an increase of 0.7% in the use of cars, and the ICLV predicts an increase of 5.2% instead; both NN and ICLV foresee a decrease in the use of *walking* for daily trips (-1.5% and -1.3% respectively), but even for this last variable we analysed, the models show contradictory result for the predictions regarding the *PT* alternative, with the NN showing an increase of 0.9% and the ICLV model saying there would be a decrease of -4.0%.

Overall, these results do not allow us to affirm with confidence that our NN model is performing correctly. While it generally gives similar results to the ICLV model when observing those relative to the *walking* alternative, the same cannot be said for *car* and *PT*. For the first one, most results of the neural network are much lower in value than those we obtained with the ICLV, and for the second they also show different signs, thus picturing a very different phenomenon. In this case, unlike we could for Model 1, since the values are significantly distant from zero, we cannot conscientiously use the justification of their value fluctuating.

On the other side, it is also possible that the ICLV model could be overestimating some of the results, since some of them assume quite extreme values (*e.g.*, those for the variable *owns a car*). But, without any other elements to give proper justifications for our results, we have to assume that our neural network in not working completely correctly when using the latent variable LV2 associated with the intentions to use sustainable means of transport.

4.5.3 Model 3 - LV1 (PBC) and LV2 (Intentions)

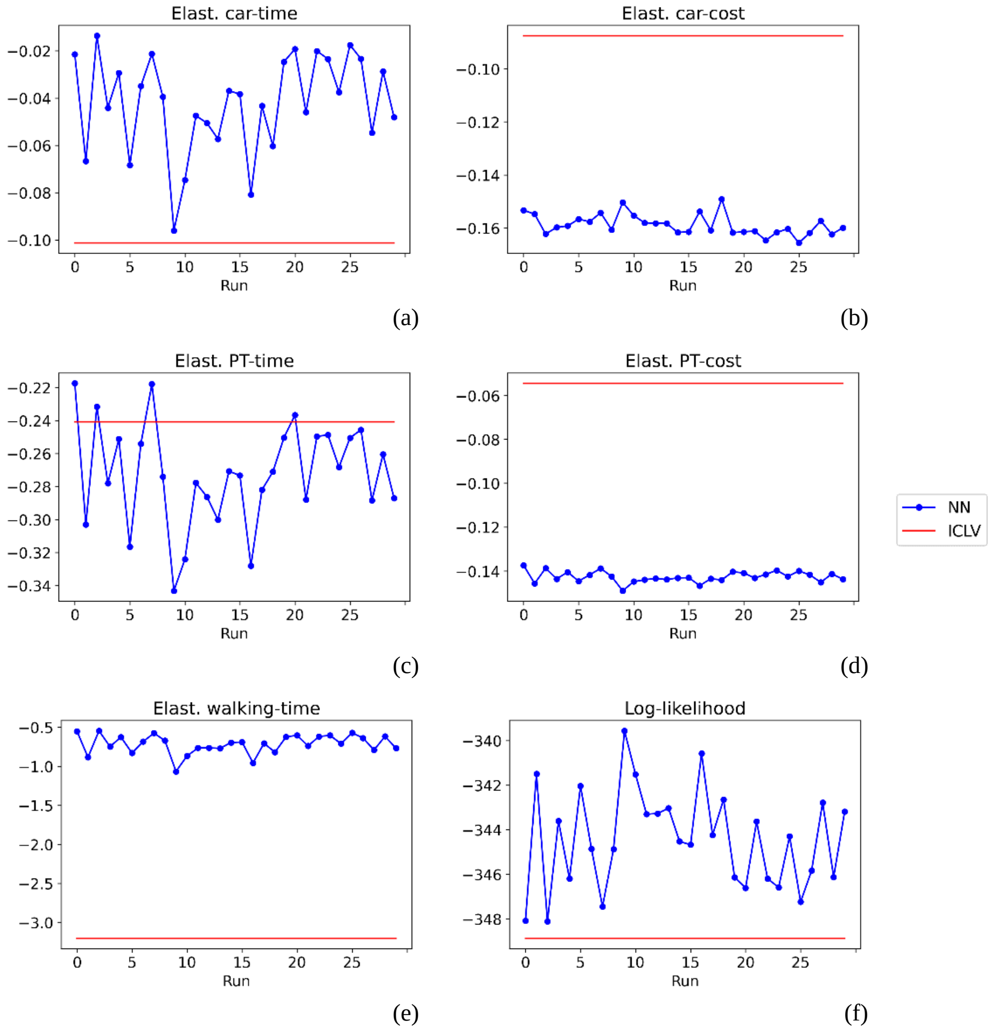
Table 25 shows all the direct- and cross-elasticities generated with Model 3 (both NN and ICLV) by considering variations in the level of services variables. Like it happened with Model 2, even if the values we obtained are slightly different, they follow the same patterns we observed in Model 1 (Table 21), meaning all the same comments are valid once again.

Table 25. Elasticities for level of service variables with Model 3 (LV1 + LV2)

Variable	Model	Car	PT	Walking
Car cost	NN	-0.161	0.233	0.024
	ICLV	-0.102	0.170	0.046
Car travel time	NN	-0.041	0.062	0.011
	ICLV	-0.122	0.210	0.101
PT cost	NN	0.080	-0.139	0.035
	ICLV	0.036	-0.059	0.066

PT travel time	**NN**	0.153	-0.282	0.057
	ICLV	0.206	-0.354	0.339
Walking travel time	**NN**	0.090	0.128	-0.722
	ICLV	0.251	0.661	-3.254

Figure 9 shows instead the pattern of the benchmark indicators we obtained with the neural network which considers both LV1 (PBC) and LV2 (Intentions), compared to those deriving from the corresponding ICLV model. The elasticities follow the same trends we already saw in both Model 1 (Paragraph 4.5.1) and Model 2 (Paragraph 4.5.2), so no further comments are needed. The log-likelihood for all NNs was higher than the one we got with the ICLV model, meaning we reached a higher level of fit.



(a) elasticity of car choice probability w.r.t. car travel time; (b) elasticity of car choice probability w.r.t. car cost; (c) elasticity of PT choice probability w.r.t. PT travel time; (d) elasticity of PT choice probability w.r.t. PT cost; (e) elasticity of walking choice probability w.r.t. walking travel time; (f) log-likelihood

Figure 9. Neural network results benchmark for Model 3 (LV1 + LV2)

Table 26 shows the pseudo-elasticities which describe the effects of variations of the latent variable LV1 (PBC) on the choice probabilities of the three alternatives, obtained both with our neural network models and the ICLV model. Like we did for Model 1 already, we will proceed by analysing the pseudo-elasticities associated with each socio-economic variable we used to build LV1.

- *Age*: if every person in our sample was 5 years older, the probability of choosing the *car* alternative would increase by 0.6% according to the NN and by 0.8% for the ICLV, and similarly the use of public transport would decrease by -0.7% (NN) and -0.5% (ICLV); but while the models agree on these two modes, they predict different result for *walking*, with the NN saying there would be an increase (0.2%) and the ICLV model a decrease (-0.4%);
- *Gender = male*: in the case where none of the individuals were male, the changes in the choice probability would be very low for all alternatives, according to both models; 0.3% (NN) or 0.3% (ICLV) for the alternative *car*, -0.4% (NN) or -0.1% (ICLV) for public transport, and 0.1% (NN) or -0.2% (ICLV) for *walking*; although the values for walking show different signs, they are very close to zero so the sign might not actually be significant;
- *Graduate*: if no one in the sample achieved at least an university degree, the NN predicts a decrease of -0.9% in the use of the *car*, and the ICLV model one of -1.2%; the probability of choosing *PT* for work/study trips would increase by 1.1% (NN) and 0.8% (ICLV), while walking would decrease by -0.3% according to the NN and increase by 0.6% for the ICLV, thus keeping the trend of contradictory results for the pseudo-elasticities for this mode obtained by modifying LV1;
- *Presence of children*: if no children were present in the households, the NN shows a decrease in the use of *car* equal to -0.6%, and the ICLV one of -1.0%, while the use of PT would rise by 0.8% (NN) or 0.7% (ICLV); once again, even though the values are very low, we got conflicting results for *walking*, since the NN predict a decrease in the choice probability (-0.2%) and the ICLV an increase (0.5%);
- *Owns a car*: in the case where no one in the sample owned a car, both models predict the largest variations in the choice probabilities among all the variables; as a matter of fact, for the *car* alternative, the NN says there would be a decrease of -8.7% and the ICLV one of -11.6%, while for *PT* there would be an increase of either 10.3% (NN) or 8.1% (ICLV); however, the models still disagree for the alternative *walking*, since the NN predicts a decrease (-1.6%) and the ICLV model an increase (5.0%)

- *Student*: if each individual was not a student, the probability of choosing the car for their work trips would increase by 2.0% according to the NN model and by 2.3% for the ICLV, while the use of public transport would decrease by -3.0% (NN) and -1.7% (ICLV); regarding the *walking* alternative, the NN predicts an increase of 0.9%, while the ICLV a decrease of -0.7%, thus depicting contradicting scenarios once again;

Table 26. Pseudo-elasticities for LV1 with Model 3 (LV1 + LV2)

Variable	Model	Car	PT	Walking
Age	**NN**	0.55%	-0.74%	0.19%
	ICLV	0.79%	-0.53%	-0.37%
Gender = male	**NN**	0.31%	-0.41%	0.10%
	ICLV	0.23%	-0.13%	-0.16%
Graduate	**NN**	-0.88%	1.15%	-0.27%
	ICLV	-1.17%	0.79%	0.55%
Presence of children	**NN**	-0.63%	0.79%	-0.16%
	ICLV	-0.99%	0.66%	0.49%
Owns a car	**NN**	-8.74%	10.33%	-1.59%
	ICLV	-11.62%	8.06%	4.98%
Student	**NN**	2.04%	-2.97%	0.93%
	ICLV	2.29%	-1.71%	-0.74%
Dep. Time 8:30-9:29 AM	**NN**	-0.62%	0.81%	-0.19%
	ICLV	-0.80%	0.55%	0.34%
House in Cagliari city	**NN**	3.11%	-5.69%	2.57%
	ICLV	6.39%	-3.58%	-4.50%
Nr. of cars	**NN**	2.58%	-3.53%	0.96%
	ICLV	3.46%	-2.35%	-1.58%

- *Dep. Time 8:30-9:29 AM*: in case no one left their house in the peak hour which starts at 8:30 AM, we would see a slight decrease of the use of the *car* alternative by -0.6% for the NN and -0.8% for the ICLV model; at the same time, the probabilities of using *PT* would increase by 0.85 (NN) and 0.6% (ICLV), but *walking* would be used less according to the NN (-0.2%) and more according to the ICLV model (0.3%);
- *House in Cagliari city*: if everyone lived outside of the city of Cagliari, the alternative *car* would increase according to both models, by 3.1% for the NN and by 6.4% for the ICLV model, while there would be a decrease in the use of public transport, -5.7% for the NN model and -3.6% for the ICLV; *walking* still produces contradictory results, since the NN says the probability would increase by 2.6% and the ICLV that it would decrease by -4.5% instead;

- *Nr. of cars*: in the case where the number of cars available in each household would increase by one unit, the use of the car alternative would in turn increase by 2.6% according to the NN, and by 3.5% for the ICLV model, while at the same time the probability of choosing public transport for work and study related trips would decrease by -3.5% (NN) or 2.4% (ICLV); even in this last case, the results for walking are not optimal, with the NN predicting an increase of 1.0% and the ICLV says the probability would instead decrease by -1.6%.

We can safely state that the performance of our neural network is very close to the one of the corresponding ICLV model, at least for the elasticities obtained for the *car* and *PT* alternatives. In fact, the pseudo-elasticities for the *walking* alternative seem to be representing different phenomena, which could be related to the different modelling frameworks used in the two specifications, which could theoretically lead to latent variables representing different aspects. These results are in line with those we got for Model 1 (Paragraph 4.5.1), albeit the values being relatively lower for Model 3. This is probably related to the fact that this last model also includes a second latent variable (LV2), meaning the latent effects they represent could have been split and distributed between the two latent constructs.

Table 27 finally shows the values of the pseudo-elasticities we obtained by inducing variations in the other latent variable included in Model 3, LV2 (Intentions). We have again compared the values obtained with our NN model and those given by the ICLV to analyse the changes in the choice probabilities of the three alternatives. We will now consider them based on each socio-economic variable which was appropriately modified.

- *Gender = male*: supposing that all the individuals in the sample were female, we would see a slight increase in the use of *walking*, ~0.2% for both models, but the results for the other alternatives are contradictory; as a matter of fact, the NN says the use of cars would increase by 0.3% and the ICLV says it would decrease instead by -1.2%; on the other hand, the NN says the use of public transport would decrease by 0.4%, against a predicted increase of 1.0% given by the ICLV model;
- *Presence of children*: if the households included no children among their members, the models agree in the fact that there would be a very slight increase in the use of walking, 0.1% for the NN and 0.2% for the ICLV model; however the results for *car* look far from ideal, with NN predicting an increase of 0.2% and the ICLV model saying instead that the probability would decrease by -0.7%; similarly, for the *PT* alternative, the NN says there would be a decrease of -0.3%, and the ICLV predicts an increase of 0.6%;

- *Owns a car*: if no one in the sample owned a car, the NN says there would be an increase in the use of the *car* alternative (2.9%), which seems counterintuitive, while the ICLV model predicts a decrease of -16.1%, more in line with what an expected result; the models also disagree on the probability of choosing *PT*, with the NN predict a decrease (-4.6%) and the ICLV an increase (13.4%); the models however give similar results for *walking*, the NN predicts an increase of 1.8% and the ICLV model one of 2.4%;

Table 27. Pseudo-elasticities for LV2 with Model 3 (LV1 + LV2)

Variable	Model	Car	PT	Walking
Gender = male	**NN**	0.27%	-0.42%	0.16%
	ICLV	-1.20%	1.00%	0.17%
Presence of children	**NN**	0.17%	-0.28%	0.11%
	ICLV	-0.71%	0.55%	0.18%
Owns a car	**NN**	2.86%	-4.61%	1.75%
	ICLV	-16.13%	13.38%	2.37%
Student	**NN**	-0.88%	1.29%	-0.40%
	ICLV	2.78%	-2.62%	0.23%
House in Cagliari city	**NN**	-0.90%	1.68%	-0.78%
	ICLV	3.28%	-2.62%	-0.69%
Nr. of cars	**NN**	-1.00%	1.48%	-0.48%
	ICLV	2.81%	-2.36%	-0.36%

- *Student*: in the case where everyone completed their studies already, the NN foresees a decrease of -0.9% in the probability of choosing the alternative *car*, while the ICLV predicts an increase of 2.8%; regarding public transport, the NN says there would be an increase of 1.3% and instead the ICLV model shows a decrease of -2.6%; for the alternative *walking*, the models still contradict each other, since the NN predicts a decrease in its use (-0.4%) and the ICLV model an increase of 0.2%;
- *House in Cagliari city*: if everyone lived outside of the city of Cagliari, the NN says we would notice a decrease in the use of cars, while the ICLV says there would be a 3.3% increase instead, and similarly the NN predicts an increase of 1.7% in the probability of using public transport while the ICLV model a decrease of -2.6%; although, the models give similar results for *walking*, with -0.8% for the NN and -0.7% for the ICLV;
- *Nr. of cars*: in the case where every household acquired an additional car, the NN predicts a decrease of -1.0% for the probability of choosing the *car* alternative, while the ICLV says there would be an increase of 2.8%; for *PT*, the NN predicts an increase of 1.5% and the ICLV model a decrease of -2.4%; and finally, for *walking*, the models agree and both says there

would be a decrease in the probability of choosing this mode (-0.5% for NN and -0.4% for the ICLV model).

These results are similar to those we obtained for Model 2 (Paragraph 4.5.2), even if the values are relatively lower in this case, again probably because the presence of LV1 in the same model redistributed the latent effects among the two factors. However, a major change when compared to Model 2 is the fact that the results for the alternative *car* have switched their signs compared to Model 2, further increasing the differences between the NN and the ICLV model.

Everything considered, the results we obtained for Model 3, by considering both latent variables (PBC and Intentions) are quite disappointing. While the pseudo-elasticities the NN predicted for the latent variable PBC were similar to those we obtained for Model 1, both with the NN and the ICLV model, and the ICLV model which included both latent variables, the pseudo-elasticities we obtained by inducing changes in LV2 (Intentions) were different both from those of Model 2 (NN and ICLV both) and those of the ICLV model with LV1 and LV2. Specifically, the fact that many elasticities changed sign, especially compared to the ICLV model (which we should consider as a solid reference), does not allow to state that this model is performing as expected, and thus the smaller values compared to the one we got with the ICLV are only a secondary issue in this case.

4.6 Discussion

Globally, none of the NN models presented significant issues when predicting the direct- and cross-elasticities connected to the level of service variables, meaning they are working correctly at least in the choice modelling module of the whole model. However, with the analysis of the pseudo-elasticities, the best results we got with a NN model were the ones from Model 1, if we accept the fact that some of the values relative to the alternative walking contradicted those given by the homologue ICLV (although some of them are so close to zero we could consider them negligible). Model 2 results were too different from those of the ICLV model built with LV2, and just the pseudo-elasticities linked to the alternative *walking* gave reasonable results. Finally, while Model 3 worked similarly to Model 1 when predicting the pseudo-elasticities connected to the first latent variable (PBC), when it came to the other one (Intentions) the results were even worse than the ones we obtained with Model 2. These results could be likely connected to the latent variables we chose to consider, or perhaps even to the fact that we skipped the exploratory factor analysis and used the pre-assumption that the psycho-attitudinal indicators where already correlated based on the behavioural theory on which their definition was based.

We will try to interpret some of the results of Model 1 (we will not consider the other two since they are considerably worst) from a phenomenological point of view, trying to highlight the benefits or disadvantages deriving from using a NN rather than a ICLV model, using Table 22 as a reference. Keep in mind that all results are relative to the indirect effects that the socio-economic variable achieves through the latent factor and should be independent from the direct effect same variable.

- *Age*: both models say that, based on PBC, older people would use the car more than younger ones, and we would just slightly underestimate the effects by using the NN; this effect is amplified for public transport, since the NN would predict a more (+0.5%) younger people to use it, compared to ICLV; according to the NN, older people would walk more, while the ICLV predicts the opposite, and this means that a policy aimed at a specific age-range, based on the results of the NN, could lead to the opposite outcome for the share of individuals traveling on foot, if the ICLV is correct instead;
- *Graduate*: both models agree in the fact that people that achieved an higher education use the car less than those who did not, and they give basically the same values; public transport is used more by people with lower level of education, but the difference predicted by the NN is higher (+0.9%) than the one shown by the ICLV; finally, more educated people walk more according to the NN, and less for the ICLV; if the outcomes of the ICLV model were correct, a policy based on the predictions of the NN would consider an overestimate for public transport, and the opposite effect for walking;
- *Owns a car*: car ownership has an obvious influence on car use, and this is probably why both models say that people who do not own a car also use the car less compared to those who do, and they return basically the same value; at the same time, if those same people need to travel, they would probably use public transport (especially for medium and long distances), and again both models correctly interpret the phenomenon, although the NN says that people who do not own a car use public transport 18.7% more compared to those that do own one, while the ICLV says this difference is much lower (11.7%, that is 7.0% less); the NN says people who do not own a car walk less than those who do, and the ICLV predicted instead the opposite; again, if ICLV results were to be true, a policy aimed at reducing the number of cars to shift more users towards sustainable transportation modes, would severely overestimate the influence of PBC on the use of public transport, and would be completely wrong with regards to the effects on walking;
- *House in Cagliari city*: the NN predicts that people not living in the city of Cagliari use the car less than those that do, which is in fact reasonable considering the destination of most

intercepted trips were indeed in the city itself; however, the difference between the two population groups, according to the NN is slightly more than half (-4.7%) of the value predicted by the ICLV model; at the same time people living further away would use less public transport, most of times because of its compatibility with their time constraints, and in this case the difference is higher for the NN (+4.3%) compared to the ICLV model; again, for walking the NN says living outside the city would increase the trips on foot while the ICLV models predicts a decrease, and the difference here is striking since the variation (in a sense or the other) would be of almost 5% for both models; this last results from the NN seems not plausible, since longer distances would seem to discourage walking even more, thus making the predictions of the ICLV seem more realistic; policies based on the PBC effects on choice behaviour, even omitting the possible inaccuracies for the use of car and public transport, would be very problematic when any prediction relative to walking is involved;

- *Nr. of cars*: like the effects of car ownership, the number of car available in the household has strong effects on the mode choice; as a matter of fact, people who can have access to more cars tend to travel by car more compared to those who have less, and the NN and the ICLV produce similar outcomes, even though the NN predicts a lower effect (-1.1%); at the same time, with easier access to cars, less people would use public transport, and in this case the NN showed that the difference between those who have more cars available and those who have less is 2.2% higher compared to the results of the ICLV; once again, results for walking are basically opposite, with the NN saying more cars equals to more trips by walking, and the ICLV says they would be less; policy aimed at reducing the number.

5 CONCLUSIONS

In this thesis we tried to highlight the benefits of using machine learning models as an alternative to discrete choice models when studying choice behaviour in transportation settings. We did so by constructing and employing different neural network architectures and comparing their results to equivalent discrete choice models.

At first, an up-to-date literature review of the applications of ML to choice modelling in transportation was used to identify some critical issue which could be addressed. This review highlighted how only a relatively small number of studies tried obtained interpretable economic information from ML models, but also how even less consider psycho-attitudinal indicators when building latent variables to represent psychological factors influencing individuals' choice.

In Chapter 3 we presented a method for obtaining the value of travel time valid for both DCs and ML models. This method was necessary since VTT is considered one of the most important indices that can be inferred, and, so far, very few researchers studied this argument in relation to ML methods. We compared the results of a neural network, specifically designed to consider separately the utilities of the different choice alternatives, with those of a multinomial logit and a mixed logit.

The first results, obtained with the *Swissmetro* dataset, showed very close results when comparing direct and cross-elasticities with respect to the level of services variables of the alternatives. We then compared the VTTs we obtained for each alternative: for the alternative *train*, the VTT obtained through the neural network (30.19 CHF/h) is almost the same of the one resulting from the MNL model (30.78 CHF/h), while the one from the MXL model (86.06 CHF/h) is quite different from both. A similar pattern was presented for the alternative *Swissmetro*, even if in this case the value obtained with the MXL (50.12 CHF/h), is closer to those of NN (40.37 CHF/h) and MNL (40.87 CHF/h). Finally, for the alternative *car*, the VTT obtained by using the NN (114.43 CHF/h) was valued between the one of MNL (90.61 CHF/h) and the one of MXL (134.46 CHF/h).

When analysing the results obtained with the *Svolta Cagliari* dataset, once again no discrepancies were found in the elasticities for the level of service variables. The VTTs for the alternative car are very similar among all the models, the NN predicted the lowest overall value (14.50 €/h), followed by the one from MNL (16.92 €/h), with MXL providing the highest value (19.67 €/h). For the alternative *PT* the VTT shows values even closer to each other, even though here the lowest value was the one obtained with the MXL (7.33 €/h), followed by the one from the NN (7.69 €/h) and finally the highest was produced by the MNL (8.25 €/h).

Considering these results, the neural network returned similar values to those of a multinomial logit, while in general there were more noticeable differences with the mixed logit results. Since there were also no problems in the elasticities, and in particular their signs were consistent with the ones suggested by the microeconomic theory, we can confidently state that this NN model is able to mimic the behaviour of a MNL model and is thus a valid alternative to discrete choice models. These results are also confirmed by the fact that we obtained consistent results with two very different datasets, from different territorial contexts (Switzerland and Italy), different years (1998 and 2019), and different survey methods (stated preferences and revealed preferences). Since these results prove that econometric indicators can be safely extracted from the outputs of a NN models, a carefully constructed neural network model could be used in the study of choice model whenever researchers might feel uncertain about the definition of the utilities for the alternatives and would prefer having the model itself choosing how to combine the variables in the data.

In Chapter 4, we proposed an alternative ML method, as an alternative to DCMs, for the specification and estimation of choice models with latent variables built around psycho-attitudinal indicators. Since several studies in social sciences have shown that choice behaviour is influenced by psychological factors, and only few studies on ML models have included latent variables, among which only a few consider psycho-attitudinal variables obtained with appropriate surveys, we felt the need to improve the knowledge in this field of study. We compared the results of a novel architecture of neural network which follows the structure of an integrate choice latent variables (ICLV) model with those of a classical ICVL model.

Three different models were estimated, by considering different latent variables: the first one (Model 1) only considered a latent variable relative to perceived behavioural control (PBC) of using sustainable alternatives; the second one (Model 2) only considered the latent variable relative to the intentions of using sustainable modes; the last one (Model 3) considered both latent variables. All results were estimated using the *Svolta Cagliari* dataset which we already used in the VTT estimation, only this time we also included some psycho-attitudinal indicators which were not needed in the previous study.

For Model 1, the neural network performs comparably to the ICLV, with the exception for the *walking* alternative, for which the NN seem to be interpreting the inverse phenomenon, even if some of the interest values are very close to zero, thus the contradiction could be due to a fluctuation of the values.

In Model 2, we have similar results to the ICLV model when observing the pseudo-elasticities relative to walking, but the same is not true for neither car nor public transport. For the first one, most results of the NN are much lower in value than those of the ICLV, and for the second alternative they also show different signs, thus the two models are picturing very different phenomena. Finally, with Model 3, we obtained values similar to those of Model 1 when considering the *PBC* latent variable, albeit the values being relatively lower (probably for the presence of interactions with the other latent variable), but for the second latent variable (*Int*) we got worse results than the ones of Model 2, since the results for the alternative *car* switched their signs, further increasing the differences between the NN and the ICLV model.

Unlike what happens when we analyse elasticities related to level of service variables, for which the microeconomic theory can help us at least recognise when a sign is incorrect, there are no such theoretical bases to rely on. We cannot thus affirm which of the two models is closer to the real phenomenon, and we can only assume that the differences between the results of the two approaches could be connected to the fact that, even though both models were constructed following the same logic, the latent variables could end up representing different aspects and should be interpreted differently, making it difficult to compare them.

If one was to use the NN results of the best of the 3 models (Model 1) as a base for policies based on perceived behavioural control, while for the most part the outcomes would be in agreement with the ICLV model, some of them would see different impacts than expected. This is especially true for walking, since often NN and ICLV disagree on the direction of the change when a socio-economic variable value is altered. For example, a striking result is the one predicted by the NN model for the effects of *House in Cagliari city*, since the NN predicts that people living outside the city have a higher probability to actively commute than those that do, which seems not plausible, since most destination of the trips are in the city of Cagliari, and longer distances should discourage walking even more.

However, the fact that the NN model and the ICLV model produced different results, could also be interpreted as a positive outcome of the study, if used correctly. As a matter of fact, this issue could lead an analyst to reflect on the advantages of using such different modelling frameworks, instead of blindly trusting one of them (*e.g.*, only the ICLV). In fact, if we built only one model and were to follow its results to build a transport policy, in the case when these results were unreliable, we would be obtaining unexpected results, potentially wasting considerable amounts of public resources. Using different models instead, would highlight those aspects for which the results are consistent among the different framework results, thus reinforcing the validity of transport policies based on them.

If we were to limit our decisions to the best model we identified, *i.e.*, Model 1 which uses only the *PBC* latent variable, and if we also considered only those variables which the NN and the ICLV model agreed upon, we could probably define some solid policy implications. Keeping in mind that the results obtained represent the latent effects of the socio-economic characteristics, rather than their direct effects on the choice probabilities, we could influence the Perceived Behavioural Control towards more sustainable means of transport by intervening on those aspects which seemed to have the most significant effects in term of elasticities.

Unsurprisingly, some of the highest elasticities were obtained for variables connected to car availability (both *Owns a car* and *Nr. of cars*). An intervention aimed at reducing the number of cars available in each household would significantly shift travel behaviour from car-use towards public transport, but at the same time these results are hardly achievable by simply implementing transport policies, even if these include hard and structural measures, some even at a national level (*e.g.*, reducing the number parking spaces in the city, implementing congestion charges, removing financial incentives for the purchase of new vehicles, increasing taxation on vehicle ownership).

After these variables, the second most prominent effect was given by the variable *House in Cagliari city*, which is could also be seen as a proxy for the distance between home and workplace. In this case, policies could be more easily implemented on two different fronts. First, public transport services could be heavily improved to bring their level of service closer to the one perceived by car users, mainly by increasing its frequency and capillarity. On the other hand, we could think of reducing the distance of the workspaces, whenever possible, by instituting satellite offices in the several municipalities surrounding the cities, or by incentivizing working from home, on a voluntary basis for the workers, at least for a few days every week. This would severely reduce the use of cars, possibly in favour of active modes like walking and cycling.

Another variable which produced higher values for the elasticities compared to the others is *Student*, which could be also linked to the students' younger age and lower financial resources. In this case, policies should act more on an educational level, and should be implemented in collaboration with high schools and universities. Since students are at a point in their lives where they still do not rely too much on the use of a car for their travels, it would be easier to convince them that it is possible to continue to use more sustainable alternatives even further in their lives. Of course, this would only be possible in an environment where the PT system is highly efficient, and if this is not true, these policies would need to be also accompanied by policies aimed at improving PT in a short time period.

While the results presented in this thesis are promising, they are still sub-optimal, especially for the latent variable NN models, so we believe there is room for improvement in future research. First of all, new and possibly better results could be obtained by using different datasets. While this is quite easy for the VTT estimation models, it is more difficult for the latent variable aspect, since datasets containing psycho-attitudinal variables are harder to come by, considering surveys containing such questions are rarer since they tend to be more expensive. Another possible research outlet would be that of modifying the structure of the NN models, to either resemble other hybrid choice models or to be independent completely and define new modelling frameworks altogether.

A limitation of this study could also be identified in the way we decided to choose which variables to include in the models, which was based on the results of a MNL for the VTT study and an ICLV for the psycho-attitudinal one. While this allowed to observe optimal results for the econometrics models, it could have led to sub-optimal results for the neural networks. A future research outlet could see different ways of selecting the variables, one of which could be the choice to not choose at all, using all the available non-correlated attributes, and disregarding the statistical significance of the parameters obtained from the logit models. Another possibility would be that of using an independent method to identify the most important variables, for example using a decision tree or a random forest algorithm to obtain the feature importance associated with each variable.

Lastly, we could also improve the selection of the hyperparameters of the neural network, which we decided to limit to a relative low number of values in order to contain the amount of time needed for the computation. This aspect could be improved by conducting a larger number of experiments, by both expanding the maximum range of values and considering smaller increments to obtain a finer distribution, but also by increasing the total number of tests for each hyperparameter.

References

1. McFadden D (1974) The measurement of urban travel demand. Journal of Public Economics 3:303–328. https://doi.org/10.1016/0047-2727(74)90003-6

2. Train K (2003) Discrete choice methods with simulation. Cambridge University Press, New York

3. Manski CF, McFadden D (1981) Structural analysis of discrete data with econometric applications. MIT press, Cambridge

4. Hausman JA, Wise DA (1978) A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences. Econometrica 46:403–426. https://doi.org/10.2307/1913909

5. McFadden D, Train K (2000) Mixed MNL models for discrete response. Journal of Applied Econometrics 15:447–470. https://doi.org/10.1002/1099-1255(200009/10)15:5<447::AID-JAE570>3.0.CO;2-1

6. Vij A, Walker JL (2016) How, when and why integrated choice and latent variable models are latently useful. Transportation Research Part B: Methodological 90:192–217. https://doi.org/10.1016/j.trb.2016.04.021

7. Ben-Akiva M, McFadden D, Gärling T, Gopinath D, Walker J, Bolduc D, Börsch-Supan A, Delquié P, Larichev O, Morikawa T, Polydoropoulou A, Rao V (1999) Extended Framework for Modeling Choice Behavior. Marketing Letters 10:187–203. https://doi.org/10.1023/A:1008046730291

8. McFadden D (1986) The Choice Theory Approach to Market Research. Marketing Science 5:275–297. https://doi.org/10.1287/mksc.5.4.275

9. Train KE, McFadden DL, Goett AA (1987) Consumer Attitudes and Voluntary Rate Schedules for Public Utilities. The Review of Economics and Statistics 69:383–391. https://doi.org/10.2307/1925525

10. Ben-Akiva M, Mcfadden D, Train K, Walker J, Bhat C, Bierlaire M, Bolduc D, Boersch-Supan A, Brownstone D, Bunch DS, Daly A, De Palma A, Gopinath D, Karlstrom A, Munizaga MA (2002) Hybrid Choice Models: Progress and Challenges. Marketing Letters 13:163–175. https://doi.org/10.1023/A:1020254301302

11. Ortelli N, Hillel T, Pereira FC, de Lapparent M, Bierlaire M (2021) Assisted specification of discrete choice models. Journal of Choice Modelling 39:100285. https://doi.org/10.1016/j.jocm.2021.100285

12. Hillel T, Bierlaire M, Elshafie MZEB, Jin Y (2021) A systematic review of machine learning classification methodologies for modelling passenger mode choice. Journal of Choice Modelling 38:100221. https://doi.org/10.1016/j.jocm.2020.100221

13. Samuel AL (1959) Some Studies in Machine Learning Using the Game of Checkers. IBM Journal of research and development 3.3:210–229

14. McCulloch WS, Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics 5:115–133. https://doi.org/10.1007/BF02478259

15. TURING AM (1950) I.—COMPUTING MACHINERY AND INTELLIGENCE. Mind LIX:433–460. https://doi.org/10.1093/mind/LIX.236.433

16. Crevier D (1993) AI: the tumultuous history of the search for artificial intelligence. Basic Books, Inc.

17. Mitchell TM (1997) Machine learning. McGraw-hill New York

18. Cunningham P, Cord M, Delany SJ (2008) Supervised Learning. In: Cord M, Cunningham P (eds) Machine Learning Techniques for Multimedia: Case Studies on Organization and Retrieval. Springer, Berlin, Heidelberg, pp 21–49

19. Greene D, Cunningham P, Mayer R (2008) Unsupervised Learning and Clustering. In: Cord M, Cunningham P (eds) Machine Learning Techniques for Multimedia: Case Studies on Organization and Retrieval. Springer, Berlin, Heidelberg, pp 51–90

20. Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. PNAS 79:2554–2558. https://doi.org/10.1073/pnas.79.8.2554

21. Svozil D, Kvasnicka V, Pospichal J (1997) Introduction to multi-layer feed-forward neural networks. Chemometrics and Intelligent Laboratory Systems 39:43–62. https://doi.org/10.1016/S0169-7439(97)00061-0

22. Likas A, Vlassis N, J. Verbeek J (2003) The global k-means clustering algorithm. Pattern Recognition 36:451–461. https://doi.org/10.1016/S0031-3203(02)00060-2

23. Bayes T (1764) An Essay towards solving a Problem in the Doctrine of Chances. Philosophical Transactions of the Royal Society of London 53:370–418

24. Rish I (2001) An empirical study of the naive Bayes classifier. IJCAI 2001 workshop on empirical methods in artificial intelligence 3:41–46

25. Breiman L, Friedman J, Stone CJ, Olshen RA (1984) Classification and Regression Trees. Taylor & Francis

26. Cortes C, Vapnik V (1995) Support-vector networks. Machine Learning 20:273–297. https://doi.org/10.1007/BF00994018

27. Doshi-Velez F, Kim B (2017) Towards A Rigorous Science of Interpretable Machine Learning

28. Michalski RS, Carbonell JG, Mitchell TM (1983) Machine Learning - An Artificial Intelligence Approach. Elsevier

29. Wang S, Mo B, Zhao J (2020) Deep neural networks for choice analysis: Architecture design with alternative-specific utility functions. Transportation Research Part C: Emerging Technologies 112:234–251. https://doi.org/10.1016/j.trc.2020.01.012

30. Wang S, Wang Q, Zhao J (2020) Deep neural networks for choice analysis: Extracting complete economic information for interpretation. Transportation Research Part C: Emerging Technologies 118:102701. https://doi.org/10.1016/j.trc.2020.102701

31. Zhao X, Yan X, Yu A, Van Hentenryck P (2020) Prediction and behavioral analysis of travel mode choice: A comparison of machine learning and logit models. Travel Behaviour and Society 20:22–35. https://doi.org/10.1016/j.tbs.2020.02.003

32. Arkoudi I, Azevedo CL, Pereira FC (2021) Combining Discrete Choice Models and Neural Networks through Embeddings: Formulation, Interpretability and Performance

33. Wong M, Farooq B (2021) ResLogit: A residual neural network logit model for data-driven choice modelling. Transportation Research Part C: Emerging Technologies 126:103050. https://doi.org/10.1016/j.trc.2021.103050

34. Han Y, Pereira FC, Ben-Akiva M, Zegras C (2022) A Neural-embedded Choice Model: TasteNet-MNL Modeling Taste Heterogeneity with Flexibility and Interpretability

35. Sfeir G, Rodrigues F, Abou-Zeid M (2022) Gaussian process latent class choice models. Transportation Research Part C: Emerging Technologies 136:103552. https://doi.org/10.1016/j.trc.2022.103552

36. Sharma S, Sharma S, Scholar U, Athaiya A (2020) Activation Functions in Neural Networks. International Journal of Engineering Applied Sciences and Technology 4:7

37. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: A Simple Way to Prevent Neural Networks from Overfitting. The journal of machine learning research 15 1929–1958

38. Friedman N, Geiger D, Goldszmidt M (1997) Bayesian Network Classifiers. Machine Learning 29:131–163. https://doi.org/10.1023/A:1007465528199

39. Macqueen J (1967) Some Methods for Classification and Analysis of Multivariate Observations. In: Proceedings of the fifth Berkeley symposium on mathematical statistics and probability. pp 281–297

40. Johnson SC (1967) Hierarchical clustering schemes. Psychometrika 32:241–254. https://doi.org/10.1007/BF02289588

41. Olson CF (1995) Parallel algorithms for hierarchical clustering. Parallel Computing 21:1313–1325. https://doi.org/10.1016/0167-8191(95)00017-I

42. Guyon I, Boser B, Vapnik V (1993) Automatic Capacity Tuning of Very Large VC-dimension Classifiers. Advances in neural information processing systems 147–155

43. Sewell M (2011) Ensemble Learning. UCL Department of Computer Science

44. Breiman L (1996) Bagging predictors. Mach Learn 24:123–140. https://doi.org/10.1007/BF00058655

45. Breiman L (1999) Random Forests - Random Features. UC Berkeley TR567 29

46. Freund Y (2001) An adaptive version of the boost by majority algorithm. Machine Learning 43:293–318. https://doi.org/10.1023/A:1010852229904

47. Friedman JH (2001) Greedy function approximation: A gradient boosting machine. Ann Statist 29:. https://doi.org/10.1214/aos/1013203451

48. Trunfio GA (2020) Recent Trends in Modelling and Simulation with Machine Learning. In: 2020 28th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP). IEEE, Västerås, Sweden, pp 352–359

49. Karpatne A, Atluri G, Faghmous JH, Steinbach M, Banerjee A, Ganguly A, Shekhar S, Samatova N, Kumar V (2017) Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data. IEEE Trans Knowl Data Eng 29:2318–2331. https://doi.org/10.1109/TKDE.2017.2720168

50. Domencich TA, McFadden D (1975) Urban Travel Demand - A Behavioral Analysis

51. McFadden D (1978) Modeling the Choice of Residential Location. Spatial Interaction Theory and Planning Models 75–96

52. Daganzo C (2014) Multinomial Probit: The Theory and Its Application to Demand Forecasting. Elsevier

53. Paleti R (2018) Generalized multinomial probit Model: Accommodating constrained random parameters. Transportation Research Part B: Methodological 118:248–262. https://doi.org/10.1016/j.trb.2018.10.019

54. Bhat CR (1995) A heteroscedastic extreme value model of intercity travel mode choice. Transportation Research Part B: Methodological 29:471–483. https://doi.org/10.1016/0191-2615(95)00015-6

55. McKelvey RD, Zavoina W (1975) A statistical model for the analysis of ordinal level dependent variables. The Journal of Mathematical Sociology 4:103–120. https://doi.org/10.1080/0022250X.1975.9989847

56. Greene WH, Hensher DA (2010) Modeling Ordered Choices: A Primer. Cambridge University Press

57. Srinivasan S, Bhat CR (2005) Modeling household interactions in daily in-home and out-of-home maintenance activity participation. Transportation 32:523–544. https://doi.org/10.1007/s11116-005-5329-z

58. Ferdous N, Eluru N, Bhat CR, Meloni I (2010) A multivariate ordered-response model system for adults' weekday activity episode generation by activity purpose and social context. Transportation Research Part B: Methodological 44:922–943. https://doi.org/10.1016/j.trb.2010.02.002

59. Bhat CR (2015) A new generalized heterogeneous data model (GHDM) to jointly model mixed types of dependent variables. Transportation Research Part B: Methodological 79:50–77. https://doi.org/10.1016/j.trb.2015.05.017

60. Van Middelkoop M, Borgers A, Timmermans H (2003) Inducing Heuristic Principles of Tourist Choice of Travel Mode: A Rule-Based Approach. Journal of Travel Research 42:75–83. https://doi.org/10.1177/0047287503254116

61. Zhang Y, Xie Y (2008) Travel Mode Choice Modeling with Support Vector Machines. Transportation Research Record 2076:141–150. https://doi.org/10.3141/2076-16

62. Zenina N, Borisov A (2011) Transportation Mode Choice Analysis Based on Classification Methods. Scientific Journal of Riga Technical University Computer Sciences 45:49–53. https://doi.org/10.2478/v10143-011-0041-2

63. Ma T-Y (2015) Bayesian Networks for Multimodal Mode Choice Behavior Modelling: A Case Study for the Cross Border Workers of Luxembourg. Transportation Research Procedia 10:870–880. https://doi.org/10.1016/j.trpro.2015.09.040

64. Pitombo CS, Salgueiro AR, da Costa ASG, Isler CA (2015) A two-step method for mode choice estimation with socioeconomic and spatial information. Spatial Statistics 11:45–64. https://doi.org/10.1016/j.spasta.2014.12.002

65. Tang L, Xiong C, Zhang L (2015) Decision tree method for modeling travel mode switching in a dynamic behavioral process. Transportation Planning and Technology 38:833–850. https://doi.org/10.1080/03081060.2015.1079385

66. Semanjski I, Lopez AJ, Gautama S (2016) Forecasting Transport Mode Use with Support Vector Machines Based Approach. ToMS 5:111–120. https://doi.org/10.7225/toms.v05.n02.002

67. Brathwaite T, Vij A, Walker JL (2017) Machine Learning Meets Microeconomics: The Case of Decision Trees and Discrete Choice. arXiv:171104826 [stat]

68. Ma T-Y, Chow JYJ, Xu J (2017) Causal structure learning for travel mode choice using structural restrictions and model averaging algorithm. Transportmetrica A: Transport Science 13:299–325. https://doi.org/10.1080/23249935.2016.1265019

69. Ding C, Cao X, Wang Y (2018) Synergistic effects of the built environment and commuting programs on commute mode choice. Transportation Research Part A: Policy and Practice 118:104–118. https://doi.org/10.1016/j.tra.2018.08.041

70. Zhu Z, Chen X, Xiong C, Zhang L (2018) A mixed Bayesian network for two-dimensional decision modeling of departure time and mode choice. Transportation 45:1499–1522. https://doi.org/10.1007/s11116-017-9770-6

71. Jing Y, Liu Y, Zhang Z, Su Y (2019) Passenger travel behaviour on Chinese high-speed railways using machine learning based on revealed-preference data. Expert Systems 36:e12422. https://doi.org/10.1111/exsy.12422

72. Pirra M, Diana M (2019) A study of tour-based mode choice based on a Support Vector Machine classifier. Transportation Planning and Technology 42:23–36. https://doi.org/10.1080/03081060.2018.1541280

73. Perez RA, Pietrzyk MC (1995) A Primer on Neural Networks in Transportation Concepts and Applications. CUTR Research Reports 50

74. Raju KA, Sikdar PK, Dhingra SL (1996) Modelling mode choice by means of an artificial neural network. Environ Plann B 23:677–683. https://doi.org/10.1068/b230677

75. Subba Rao PV, Sikdar PK, Krishna Rao KV, Dhingra SL (1998) Another insight into artificial neural networks through behavioural analysis of access mode choice. Computers, Environment and Urban Systems 22:485–496. https://doi.org/10.1016/S0198-9715(98)00036-2

76. Hensher DA, Ton TT (2000) A comparison of the predictive potential of artificial neural networks and nested logit models for commuter mode choice. 18

77. Cantarella GE, de Luca S (2003) Modeling transportation mode choice through artificial neural networks. In: Fourth International Symposium on Uncertainty Modeling and Analysis, 2003. ISUMA 2003. IEEE, College Park, Maryland, USA, pp 84–90

78. Cantarella GE, de Luca S (2005) Multilayer feedforward networks for transportation mode choice analysis: An analysis and a comparison with random utility models. Transportation Research Part C: Emerging Technologies 13:121–155. https://doi.org/10.1016/j.trc.2005.04.002

79. Chalumuri RS, Errampalli M, Bujangan KD, Subamay G (2009) Applications of Neural Networks in Mode Choice Modelling for Second Order Metropolitan Cities of India. Proceedings of the Eastern Asia Society for Transportation Studies 16

80. Zhao D, Shao C, Li J, Dong C, Liu Y (2010) Travel Mode Choice Modeling Based on Improved Probabilistic Neural Network. In: Traffic and Transportation Studies 2010. American Society of Civil Engineers, Kunming, China, pp 685–695

81. Yin H, Guan H (2011) Traffic mode choice model based on BP neural network. In: Proceedings 2011 International Conference on Transportation, Mechanical, and Electrical Engineering (TMEE). IEEE, ChangChun, China, pp 1441–1444

82. Gao J, Zhao P, Zhuge C, Zhang H, McCormack ED (2013) Impact of Transit Network Layout on Resident Mode Choice. Mathematical Problems in Engineering 2013:1–8. https://doi.org/10.1155/2013/452735

83. Omrani H, Charif O, Gerber P, Awasthi A, Trigano P (2013) Prediction of Individual Travel Mode with Evidential Neural Network Model. Transportation Research Record 2399:1–8. https://doi.org/10.3141/2399-01

84. Omrani H (2015) Predicting Travel Mode of Individuals by Machine Learning. Transportation Research Procedia 10:840–849. https://doi.org/10.1016/j.trpro.2015.09.037

85. Hussain HD, Mohammed AM, Salman AD, Borhan MN (2017) Analysis of Transportation Mode Choice Using a Comparison of Artificial Neural Network and Multinomial Logit Models. 12:11

86. Assi KJ, Nahiduzzaman KM, Ratrout NT, Aldosary AS (2018) Mode choice behavior of high school goers: Evaluating logistic regression and MLP neural networks. Case Studies on Transport Policy 6:225–230. https://doi.org/10.1016/j.cstp.2018.04.006

87. Golshani N, Shabanpour R, Mahmoudifard SM, Derrible S, Mohammadian A (2018) Modeling travel mode and timing decisions: Comparison of artificial neural networks and copula-based joint model. Travel Behaviour and Society 10:21–32. https://doi.org/10.1016/j.tbs.2017.09.003

88. Lee D, Derrible S, Pereira FC (2018) Comparison of Four Types of Artificial Neural Network and a Multinomial Logit Model for Travel Mode Choice Modeling. Transportation Research Record 2672:101–112. https://doi.org/10.1177/0361198118796971

89. Minal S, Sekhar ChR (2018) Web Survey Data and Commuter Mode Choice Analysis Using Artificial Neural Network. IJTTE 8:359–371. https://doi.org/10.7708/ijtte.2018.8(3).08

90. Aschwanden GD, Wijnands JS, Thompson J, Nice KA, Zhao H, Stevenson M (2019) Learning to walk: Modeling transportation mode choice distribution through neural networks. Environment and Planning B: Urban Analytics and City Science 239980831986257. https://doi.org/10.1177/2399808319862571

91. Ma Y, Zhang Z (2020) Travel Mode Choice Prediction Using Deep Neural Networks With Entity Embeddings. IEEE Access 8:64959–64970. https://doi.org/10.1109/ACCESS.2020.2985542

92. Zhang Z, Ji C, Wang Y, Yang Y (2020) A Customized Deep Neural Network Approach to Investigate Travel Mode Choice with Interpretable Utility Information. Journal of Advanced Transportation 2020:1–11. https://doi.org/10.1155/2020/5364252

93. Buijs R, Koch T, Dugundji E (2021) Using neural nets to predict transportation mode choice: Amsterdam network change analysis. J Ambient Intell Human Comput 12:121–135. https://doi.org/10.1007/s12652-020-02855-6

94. Li X, Wang Y, Wu Y, Chen J, Zhou J (2021) Modeling Intercity Travel Mode Choice with Data Balance Changes: A Comparative Analysis of Bayesian Logit Model and Artificial Neural Networks. Journal of Advanced Transportation 2021:e9219176. https://doi.org/10.1155/2021/9219176

95. van Cranenburgh S, Kouwenhoven M (2021) An artificial neural network based method to uncover the value-of-travel-time distribution. Transportation 48:2545–2583. https://doi.org/10.1007/s11116-020-10139-3

96. Biagioni JP, Szczurek PM, Nelson PC, Mohammadian A (2008) Tour-Based Mode Choice Modeling: Using An Ensemble of (Un-) Conditional Data-Mining Classifiers. 16

97. Rasouli S, Timmermans HJP (2014) Using ensembles of decision trees to predict transport mode choice decisions: Effects on predictive success and uncertainty estimates. 13

98. Ermagun A, Rashidi TH, Lari ZA (2015) Mode Choice for School Trips: Long-Term Planning and Impact of Modal Specification on Policy Assessments. Transportation Research Record 2513:97–105. https://doi.org/10.3141/2513-12

99. Sekhar ChR, Minal S, Madhu E (2016) Mode Choice Analysis Using Random Forrest Decision Trees. Transportation Research Procedia 17:644–652. https://doi.org/10.1016/j.trpro.2016.11.119

100. Brondeel R, Kestens Y, Chaix B (2017) An evaluation of transport mode shift policies on transport-related physical activity through simulations based on random forests. Int J Behav Nutr Phys Act 14:143. https://doi.org/10.1186/s12966-017-0600-1

101. Wang F, Ross CL (2018) Machine Learning Travel Mode Choices: Comparing the Performance of an Extreme Gradient Boosting Model with a Multinomial Logit Model. Transportation Research Record 2672:35–45. https://doi.org/10.1177/0361198118773556

102. Chang X, Wu J, Liu H, Yan X, Sun H, Qu Y (2019) Travel mode choice: a data fusion model using machine learning methods and evidence from travel diary survey data. Transportmetrica A: Transport Science 15:1587–1612. https://doi.org/10.1080/23249935.2019.1620380

103. Chapleau R, Gaudette P, Spurr T (2019) Application of Machine Learning to Two Large-Sample Household Travel Surveys: A Characterization of Travel Modes. Transportation Research Record 2673:173–183. https://doi.org/10.1177/0361198119839339

104. Cheng L, Chen X, De Vos J, Lai X, Witlox F (2019) Applying a random forest method approach to model travel mode choice behavior. Travel Behaviour and Society 14:1–10. https://doi.org/10.1016/j.tbs.2018.09.002

105. Lee D, Mulrow J, Haboucha CJ, Derrible S, Shiftan Y (2019) Attitudes on Autonomous Vehicle Adoption using Interpretable Gradient Boosting Machine. Transportation Research Record 2673:865–878. https://doi.org/10.1177/0361198119857953

106. Li X, Gao Y, Zhang H, Liao Y (2020) Passenger Travel Behavior in Public Transport Corridor After the Operation of Urban Rail Transit: A Random Forest Algorithm Approach. IEEE Access 8:211303–211314. https://doi.org/10.1109/ACCESS.2020.3038831

107. Ceccato R, Chicco A, Diana M (2021) Evaluating car-sharing switching rates from traditional transport means through logit models and Random Forest classifiers. Transportation Planning and Technology 1–16. https://doi.org/10.1080/03081060.2020.1868084

108. Kim E-J (2021) Analysis of Travel Mode Choice in Seoul Using an Interpretable Machine Learning Approach. Journal of Advanced Transportation 2021:e6685004. https://doi.org/10.1155/2021/6685004

109. Xie C, Lu J, Parkany E (2003) Work Travel Mode Choice Modeling with Data Mining: Decision Trees and Neural Networks. Transportation Research Record 1854:50–61. https://doi.org/10.3141/1854-06

110. Shukla N, Ma J, Denagamage RW, Huynh NN (2013) Data-driven modeling and analysis of household travel mode choice. 9

111. Hagenauer J, Helbich M (2017) A comparative study of machine learning classifiers for modeling travel mode choice. Expert Systems with Applications 78:273–282. https://doi.org/10.1016/j.eswa.2017.01.057

112. Lindner A, Pitombo CS, Cunha AL (2017) Estimating motorized travel mode choice using classifiers: An application for high-dimensional multicollinear data. Travel Behaviour and Society 6:100–109. https://doi.org/10.1016/j.tbs.2016.08.003

113. Assi KJ, Shafiullah M, Nahiduzzaman KM, Mansoor U (2019) Travel-To-School Mode Choice Modelling Employing Artificial Intelligence Techniques: A Comparative Study. Sustainability 11:4484. https://doi.org/10.3390/su11164484

114. Richards MJ, Zill JC (2019) Modelling mode choice with machine learning algorithms. 15

115. Zhao X, Yan X, Van Hentenryck P (2019) Modeling Heterogeneity in Mode-Switching Behavior Under a Mobility-on-Demand Transit System: An Interpretable Machine Learning Approach. arXiv:190202904 [cs, stat]

116. Zhou X, Wang M, Li D (2019) Bike-sharing or taxi? Modeling the choices of travel mode in Chicago using machine learning. Journal of Transport Geography 79:102479. https://doi.org/10.1016/j.jtrangeo.2019.102479

117. Aghaabbasi M, Shekari ZA, Shah MZ, Olakunle O, Armaghani DJ, Moeinaddini M (2020) Predicting the use frequency of ride-sourcing by off-campus university students through random forest and Bayesian network techniques. Transportation Research Part A: Policy and Practice 136:262–281. https://doi.org/10.1016/j.tra.2020.04.013

118. Liang L, Xu M, Grant-Muller S, Mussone L (2020) Household travel mode choice estimation with large-scale data—an empirical analysis based on mobility data in Milan. International Journal of Sustainable Transportation 15:70–85. https://doi.org/10.1080/15568318.2019.1686782

119. Koushik AN, Manoj M, Nezamuddin N (2020) Machine learning applications in activity-travel behaviour research: a review. Transport Reviews 40:288–311. https://doi.org/10.1080/01441647.2019.1704307

120. Wang S, Mo B, Hess S, Zhao J (2021) Comparing hundreds of machine learning classifiers and discrete choice models in predicting travel behavior: an empirical benchmark. arXiv preprint arXiv:210201130

121. Gazder U, Ratrout NT (2015) A new logit-artificial neural network ensemble for mode choice modeling: a case study for border transport: LOGIT-ARTIFICIAL NEURAL NETWORK ENSEMBLE FOR MODE CHOICE MODELING. J Adv Transp 49:855–866. https://doi.org/10.1002/atr.1306

122. Sifringer B, Lurkin V, Alahi A (2018) Enhancing Discrete Choice Models with Neural Networks. 15

123. Sfeir G, Abou-Zeid M, Rodrigues F, Pereira FC, Kaysi I (2021) Latent class choice model with a flexible class membership component: A mixture model approach. Journal of Choice Modelling 41:100320. https://doi.org/10.1016/j.jocm.2021.100320

124. Han Y, Pereira FC, Ben-Akiva M, Zegras C (2022) A neural-embedded discrete choice model: Learning taste representation with strengthened interpretability. Transportation Research Part B: Methodological 163:166–186. https://doi.org/10.1016/j.trb.2022.07.001

125. Small K (2013) Urban transportation economics. Taylor & Francis

126. Jara-Díaz SR (2007) Allocation and valuation of travel-time savings. In: Handbook of transport modelling. Emerald Group Publishing Limited

127. Román C, Martín JC, Espino R, Cherchi E, Ortúzar J de D, Rizzi LI, González RM, Amador FJ (2014) Valuation of travel time savings for intercity travel: The Madrid-Barcelona corridor. Transport Policy 36:105–117. https://doi.org/10.1016/j.tranpol.2014.07.007

128. Jara-Díaz S (2007) Transport Economic Theory. Emerald Group Publishing Limited

129. Bishop CM (2006) Pattern recognition and machine learning. Springer, New York

130. Hara K, Saito D, Shouno H (2015) Analysis of function of rectified linear unit used in deep learning. In: 2015 International Joint Conference on Neural Networks (IJCNN). IEEE, Killarney, Ireland, pp 1–8

131. Bierlaire M, Axhausen K, Abay G (2001) The acceptance of modal innovation: The case of Swissmetro. Proceedings of the Swiss Transport Research Conference

132. Bierlaire M (2016) PythonBiogeme: a short introduction. Report TRANSP-OR 160706, Series on Biogeme. Transport and Mobility Laboratory, School of Architecture, Civil and Environmental Engineering, Ecole Polytechnique Fédérale de Lausanne, Switzerland

133. He K, Zhang X, Ren S, Sun J (2015) Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In: 2015 IEEE International Conference on Computer Vision (ICCV). IEEE, Santiago, Chile, pp 1026–1034

134. Glorot X, Bengio Y Understanding the difficulty of training deep feedforward neural networks. 8

135. Goodfellow I, Bengio Y, Courville A (2016) Deep Learning. MIT Press

136. Dauphin YN, Pascanu R, Gulcehre C, Cho K, Ganguli S, Bengio Y (2014) Identifying and attacking the saddle point problem in high-dimensional non-convex optimization. Advances in Neural Information Processing Systems 27 (NIPS 2014) 9

137. Boyd S, Boyd SP, Vandenberghe L (2004) Convex Optimization. Cambridge University Press

138. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, Desmaison A, Köpf A, Yang EZ, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L, Bai J, Chintala S (2019) PyTorch: An Imperative Style, High-Performance Deep Learning Library. CoRR abs/1912.01703:

139. Abou-Zeid M, Ben-Akiva M (2011) The effect of social comparisons on commute well-being. Transportation Research Part A: Policy and Practice 45:345–361. https://doi.org/10.1016/j.tra.2011.01.011

140. Kamargianni M, Polydoropoulou A (2013) Hybrid Choice Model to Investigate Effects of Teenagers’ Attitudes toward Walking and Cycling on Mode Choice Behavior. Transportation Research Record 2382:151–161. https://doi.org/10.3141/2382-17

141. Vij A, Carrel A, Walker JL (2013) Incorporating the influence of latent modal preferences on travel mode choice behavior. Transportation Research Part A: Policy and Practice 54:164–178. https://doi.org/10.1016/j.tra.2013.07.008

142. Paulssen M, Temme D, Vij A, Walker JL (2014) Values, attitudes and travel behavior: a hierarchical latent variable mixed logit model of travel mode choice. Transportation 41:873–888. https://doi.org/10.1007/s11116-013-9504-3

143. Prato CG, Bekhor S, Pronello C (2012) Latent variables and route choice behavior. Transportation 39:299–319. https://doi.org/10.1007/s11116-011-9344-y

144. Daziano RA, Bolduc D (2013) Incorporating pro-environmental preferences towards green automobile technologies through a Bayesian hybrid choice model. Transportmetrica A: Transport Science 9:74–106. https://doi.org/10.1080/18128602.2010.524173

145. Glerum A, Stankovikj L, Thémans M, Bierlaire M (2013) Forecasting the Demand for Electric Vehicles: Accounting for Attitudes and Perceptions. Transportation Science. https://doi.org/10.1287/trsc.2013.0487

146. Thorhauge M, Haustein S, Cherchi E (2016) Accounting for the Theory of Planned Behaviour in departure time choice. Transportation Research Part F: Traffic Psychology and Behaviour 38:94–105. https://doi.org/10.1016/j.trf.2016.01.009

147. Piras F, Sottile E, Meloni I (2021) Do psycho-attitudinal factors vary with individuals' cycling frequency? A hybrid ordered modeling approach. Travel Behaviour and Society 22:186–198. https://doi.org/10.1016/j.tbs.2020.09.007

148. Piras F, Manca S, Sottile E, Meloni I, Mosca O, Tuveri G, Fornara F (2021) Can persuasive normative messages encourage sustainable transportation usage? Transportation Research Part F: Traffic Psychology and Behaviour 83:304–322. https://doi.org/10.1016/j.trf.2021.10.014

149. Piras F, Sottile E, Tuveri G, Meloni I (2021) Could psychosocial variables help assess pro-cycling policies? Transportation Research Part A: Policy and Practice 154:108–128. https://doi.org/10.1016/j.tra.2021.10.003

150. Wong M, Farooq B, Bilodeau G-A (2018) Discriminative conditional restricted Boltzmann machine for discrete choice and latent variable modelling. Journal of Choice Modelling 29:152–168. https://doi.org/10.1016/j.jocm.2017.11.003

151. Cao W, Mirjalili V, Raschka S (2020) Rank consistent ordinal regression for neural networks with application to age estimation. Pattern Recognition Letters 140:325–331. https://doi.org/10.1016/j.patrec.2020.11.008

152. Bhat CR (1998) A model of post home-arrival activity participation behavior. Transportation Research Part B: Methodological 32:387–400. https://doi.org/10.1016/S0191-2615(98)00006-X

153. Vij A, Walker JL (2016) How, when and why integrated choice and latent variable models are latently useful. Transportation Research Part B: Methodological 90:192–217. https://doi.org/10.1016/j.trb.2016.04.021

154. Bhat CR (1998) Analysis of travel mode and departure time choice for urban shopping trips. Transportation Research Part B: Methodological 32:361–371. https://doi.org/10.1016/S0191-2615(98)00004-6

155. De Witte A, Hollevoet J, Dobruszkes F, Hubert M, Macharis C (2013) Linking modal choice to motility: A comprehensive review. Transportation Research Part A: Policy and Practice 49:329–341. https://doi.org/10.1016/j.tra.2013.01.009

156. Halton JH (1970) A Retrospective and Prospective Survey of the Monte Carlo Method. SIAM Rev 12:1–63. https://doi.org/10.1137/1012001

157. Likert R (1932) A technique for the measurement of attitudes. Archives of Psychology 22 140:55–55

158. Manca S, Fornara F (2019) Attitude Toward Sustainable Transport as a Function of Source and Argument Reliability and Anticipated Emotions. Sustainability 11:3288. https://doi.org/10.3390/su11123288

159. Bamberg S, Hunecke M, Blöbaum A (2007) Social context, personal norms and the use of public transportation: Two field studies. Journal of Environmental Psychology 27:190–203. https://doi.org/10.1016/j.jenvp.2007.04.001

160. Klöckner CA, Friedrichsmeier T (2011) A multi-level approach to travel mode choice – How person characteristics and situation specific aspects determine car use in a student sample. Transportation Research Part F: Traffic Psychology and Behaviour 14:261–277. https://doi.org/10.1016/j.trf.2011.01.006

161. Fornara F, Carrus G, Passafaro P, Bonnes M (2011) Distinguishing the sources of normative influence on proenvironmental behaviors: The role of local norms in household waste recycling. Group Processes & Intergroup Relations 14:623–635. https://doi.org/10.1177/1368430211408149

162. Fornara F, Pattitoni P, Mura M, Strazzera E (2016) Predicting intention to improve household energy efficiency: The role of value-belief-norm theory, normative and informational influence, and specific attitude. Journal of Environmental Psychology 45:1–10. https://doi.org/10.1016/j.jenvp.2015.11.001

163. Bollen KA (1989) Structural Equations with Latent Variables. John Wiley & Sons

164. Kaiser HF, Rice J (1974) Little Jiffy, Mark Iv. Educational and Psychological Measurement 34:111–117. https://doi.org/10.1177/001316447403400115

165. Cronbach LJ (1951) Coefficient alpha and the internal structure of tests. Psychometrika 16:297–334. https://doi.org/10.1007/BF02310555

166. Nunnally JC (1967) Psychometric theory. McGraw-Hill, New York, NY, US

167. Zhao X, Yan X, Yu A, Van Hentenryck P (2019) Modeling Stated Preference for Mobility-on-Demand Transit: A Comparison of Machine Learning and Logit Models. arXiv:181101315 [cs, stat]

List of Tables

Table 1. Details for the literature regarding Bayesian Algorithms, Decision Trees, and Support Vector Machines 14

Table 2. Details for the literature regarding Neural Networks (part 1 of 2)..18

Table 3. Details for the literature regarding Neural Networks (part 2 of 2)..19

Table 4. Details for the literature regarding Ensemble Methods 22

Table 5. Details for the literature regarding Several Algorithms...25

Table 6. Details for the literature regarding “Hybrid” Methods 27

Table 7. Comparison between correct formulas and discrete approximations for elasticities and VTT 34

Table 8. Data analysis – Swissmetro database 35

Table 9. Data analysis – Svolta Cagliari database 38

Table 10. Variables used for the Swissmetro dataset with their MNL parameters and their statistical significance 39

Table 11. Variables used for the Svolta Cagliari dataset with their MNL parameters and their statistical significance 40

Table 12. Hyperparameters sets for both datasets 40

Table 13. Elasticities comparison of MNL, MXL, NN for the Swissmetro dataset...42

Table 14. Values of travel time and log-likelihood comparison of MNL, MXL, NN for the Swissmetro dataset................43

Table 15. Elasticities comparison of MNL, MXL, NN for the Svolta Cagliari dataset 46

Table 16. Values of travel time and log-likelihood comparison of MNL, MXL, NN for the Svolta Cagliari dataset 48

Table 17. Analysis of the responses to the psycho-attitudinal survey questions..59

Table 18. Confirmatory factor analysis for the psycho-attitudinal indicators..61

Table 19. Modelling results obtained with PythonBiogeme for the ICLV models...62

Table 20. Hyperparameters sets for all models..63

Table 21. Elasticities for level of service variables with Model 1 (LV1) 64

Table 22. Pseudo-elasticities for LV1 with Model 1 (LV1) 66

Table 23. Elasticities for level of service variables with Model 2 (LV2) 70

Table 24. Pseudo-elasticities for LV2 with Model 2 (LV2) 71

Table 25. Elasticities for level of service variables with Model 3 (LV1 + LV2)..72

Table 26. Pseudo-elasticities for LV1 with Model 3 (LV1 + LV2)..75

Table 27. Pseudo-elasticities for LV2 with Model 3 (LV1 + LV2)..77

LIST OF FIGURES

Figure 1. Alternative specific deep NN by Wang et al. [29]....33

Figure 2. Proposed neural network architecture. 33

Figure 3. Neural network results for the Swissmetro Dataset....44

Figure 4. Neural network results for the Svolta Cagliari Dataset 47

Figure 5. Framework for the Integrated Choice Latent Variable model by Ben-Akiva et al. [10]....53

Figure 6. Proposed neural network architecture to emulate an ICLV model. 57

Figure 7. Neural network benchmark results for Model 1 (LV1) 64

Figure 8. Neural network benchmark results for Model 2 (LV2) 69

Figure 9. Neural network results benchmark for Model 3 (LV1 + LV2)....74

Appendix A

Descriva cortesemente lo spostamento, con destinazione il proprio posto di lavoro/studio, che effettua ABITUALMENTE in un giorno feriale lavorativo

Indichi il Comune e l'indirizzo (via e numero civico) di ORIGINE del suo spostamento

Indichi il Comune e l'indirizzo (via e numero civico) della DESTINAZIONE dello spostamento

Indichi, mediamente, la frequenza con la quale si reca presso il proprio luogo di lavoro/studio
- Raramente: qualche volta all'anno (indichi nel riquadro sottostante il numero di volte)
- Talvolta: poche volte al mese (indichi nel riquadro sottostante il numero di volte)
- Spesso: più volte alla settimana (indichi nel riquadro sottostante il numero di volte)

Indichi in quale fascia oraria inizia lo spostamento
- 5:30-7:30
- 7:30-8:30
- 8:30-9:30
- 9:30-12:30
- 12:30-14:30
- 14:30-17:00
- 17:00-19:30
- 19:30-22:00

Con quale mezzo di trasporto arriva presso il proprio luogo di lavoro/studio?
- Auto, come conducente
- Auto, come passeggero
- Trasporto pubblico (autobus, metro, treno, auto + trasporto pubblico, etc.)
- Moto / ciclomotore
- Bicicletta / Bike sharing (Cabubi)
- Piedi
- Car sharing (Playcar) / Car pooling (condivisione del veicolo con colleghi)
- Altro

Indichi IN ORDINE DI PREFERENZA le tre modalità di trasporto con le quali effettuerebbe lo spostamento con motivazione lavoro/studio.
- Automobile privata
- Moto / ciclomotore
- Trasporto pubblico
- Mobilità attiva (piedi o bicicletta)
- Car sharing / Car pooling
- Auto + trasporto pubblico

Indichi il grado di accordo con le seguenti affermazioni
(1 = Fortemente in disaccordo 2 = Abbastanza in disaccordo 3 = Né in disaccordo né d'accordo
4 = Abbastanza d'accordo 5 = Fortemente d'accordo)

	1	2	3	4	5
La mia scelta di utilizzare il trasporto sostenibile è consapevolmente e INTENZIONALMENTE motivata da un desiderio specifico di far del bene all'ambiente.					
La mia scelta di utilizzare il trasporto sostenibile è INTENZIONALMENTE motivata dal fatto che è più conveniente (tempi e costi).					
La mia scelta di utilizzare il trasporto sostenibile è INTENZIONALMENTE motivata dal fatto che mi consente di fare attività fisica.					
La mia scelta di utilizzare il trasporto sostenibile è OBBLIGATA dal fatto che non ho alternative.					

Quali mezzi NON avrebbe potuto utilizzare per effettuare questo spostamento?

- Auto, come conducente
- Auto, come passeggero
- Trasporto pubblico (autobus, metro, treno, auto + trasporto pubblico, etc.)
- Moto / ciclomotore
- Bicicletta / Bike sharing (Cabubi)
- Piedi
- Car sharing (Playcar) / Car pooling (condivisione del veicolo con colleghi)

Effettua ABITUALMENTE fermate durante lo spostamento?

- Sì
- No

Indichi il luogo (Comune e via) dove avviene la fermata, o i luoghi in caso di più fermate.

Sono disponibili, dentro il luogo di lavoro/studio, parcheggi per le biciclette?

- Sì
- No

Dove parcheggia l'automobile per recarsi al luogo di lavoro/studio?

- Parcheggio interno riservato ai dipendenti
- Parcheggio gratuito su strada per automobili
- Parcheggio a pagamento su strada/struttura per automobili
- Parcheggio per moto/ciclomotori
- Garage privato

Che mezzo utilizza, alla fine dell'orario di lavoro/studio, per andare via?

- Lo stesso mezzo usato per l'andata
- Un mezzo diverso da quello usato per l'andata (specificare)

Dove si reca dopo aver lasciato il proprio luogo di lavoro/studio?

- Casa
- Altro luogo (indichi Comune e via)

Descriva l'autovettura utilizzata per il suo spostamento
Indichi il tipo di auto utilizzata per lo spostamento descritto

- City car (es. Fiat 500, Smart)
- Utilitaria (es. Fiat Punto)
- Compatta (es. VW Golf, Opel Astra)
- Station Wagon (es. Ford Focus, Fiat Tipo)
- Berlina sportiva (es. Alfa Romeo Giulia, Mercedes classe C)
- SUV
- Van

Indichi il tipo di alimentazione dell'auto

- Benzina
- Diesel
- Gpl
- Ibrida
- Elettrica

Indichi quanti km percorre mediamente all'anno in auto

Descriva le caratteristiche dello spostamento effettuato con il trasporto pubblico
Indichi il mezzo o la combinazione di mezzi utilizzati per raggiungere il luogo di lavoro/studio

- Autobus CTM
- Autobus extraurbano (CTM o ARST)
- Metrocagliari
- Treno
- Combinato bus + bus
- Combinato metro + bus
- Combinato treno + bus
- Altra combinazione

Come arriva alla fermata di SALITA del mezzo?

- Piedi
- Bicicletta
- Auto come conducente
- Auto come passeggero
- Altro

Come arriva dalla fermata di DISCESA al luogo di lavoro/studio?

- Piedi
- Bicicletta
- Auto come passeggero
- Altro

Indichi la linea di trasporto pubblico che utilizza

Indichi il tipo di biglietto che utilizza

- CTM - Biglietto ordinario a tempo (90 min)
- CTM - Biglietto multiplo da 12 corse
- CTM - Biglietto integrato a tempo (2 ore)
- CTM - Biglietto integrato giornaliero (24 ore)
- CTM - Carta integrata settimanale
- CTM - Abbonamento mensile
- CTM - Abbonamento annuale
- ARST - Corsa semplice
- ARST - Biglietto giornaliero
- ARST - Carnet 12 corse
- ARST - Abbonamento settimanale
- ARST - Abbonamento mensile
- ARST - Abbonamento annuale
- Integrato CTM + ARST + Baire
- Integrato CTM + Trenitalia
- Altro

Nelle prossime domande le verrà chiesta la sua opinione (in termini di livello di accordo o disaccordo su determinate affermazioni) riguardo le modalità di trasporto sostenibile. Con trasporto sostenibile si intendono quelle modalità di spostamento alternative all'utilizzo dell'automobile privata, come:

- *Piedi;*
- *Bicicletta / Bike sharing (Cabubi);*
- *Trasporto pubblico (autobus, tram, treno, auto + trasporto pubblico, etc.);*
- *Car sharing (Playcar) / Car pooling (condivisione del veicolo con colleghi).*

Le domande si riferiscono allo spostamento ABITUALE casa - lavoro / casa - studio.
Indichi il grado di accordo con le seguenti affermazioni
(1 = Fortemente in disaccordo 2 = Abbastanza in disaccordo 3 = Né in disaccordo né d'accordo
4 = Abbastanza d'accordo 5 = Fortemente d'accordo)

	1	2	3	4	5
La maggior parte delle persone che conosco pensano che dovrei utilizzare i mezzi di trasporto sostenibile anziché l'auto privata.					
La maggior parte delle persone che conosco utilizzano i mezzi di trasporto sostenibile invece dell'auto privata.					
Mi sento moralmente obbligato ad utilizzare i mezzi di trasporto sostenibile indipendentemente da quello che fanno gli altri.					
La mia scelta di utilizzare il trasporto sostenibile è OBBLIGATA dal fatto che non ho alternative.					
Se nelle prossime due settimane UTILIZZERO' l'auto privata e NON UTILIZZERO' il trasporto sostenibile, penso che potrei sentirmi COLPEVOLE.					
Se nelle prossime due settimane UTILIZZERO' il trasporto sostenibile e NON UTILIZZERO' l'auto privata, penso che potrei sentirmi ORGOGLIOSO.					
Se nelle prossime due settimane UTILIZZERO' l'auto privata e NON UTILIZZERO' il trasporto sostenibile, penso che potrei sentirmi INDIFFERENTE.					
Mi sento personalmente responsabile dei PROBLEMI AMBIENTALI che possono derivare dalla scelta del mio modo di trasporto.					
Mi sento personalmente responsabile dei problemi legati al TRAFFICO, all'OCCUPAZIONE DI SPAZIO PER LA SOSTA, all'INCIDENTALITA' STRADALE presenti nella mia città.					
Sono consapevole che l'utilizzo dell'automobile produce danni all'ambiente e alla salute delle persone.					
Sono consapevole che posso contribuire PERSONALMENTE (utilizzando meno l'auto per i miei spostamenti) a migliorare l'ambiente.					

Pensando alla città di Cagliari, indichi il grado di accordo con le seguenti affermazioni
(1 = Fortemente in disaccordo 2 = Abbastanza in disaccordo 3 = Né in disaccordo né d'accordo
4 = Abbastanza d'accordo 5 = Fortemente d'accordo)

	1	2	3	4	5
Questa città è parte di me e quindi la rispetto.					
Mi sento a casa in questa città.					
Mi sento completamente parte di questa città e quindi devo contribuire a renderla migliore.					

Indichi con quale frequenza ha utilizzato i seguenti mezzi di trasporto per motivazione DIVERSA dallo spostamento casa-lavoro/studio

	Mai	Qualche volta all'anno	1 - 3 volte al mese	1-4 volte a settimana	5 o più volte a settimana
Piedi					
Bici					
Trasporto pubblico					
Automobile					

Indichi il grado di accordo con le seguenti affermazioni
(1 = Fortemente in disaccordo 2 = Abbastanza in disaccordo 3 = Né in disaccordo né d'accordo
4 = Abbastanza d'accordo 5 = Fortemente d'accordo)

	1	2	3	4	5
Durante le prossime due settimane intendo utilizzare mezzi di trasporto sostenibile al posto dell'auto privata (da solo/a).					
Nelle prossime due settimane ho intenzione di utilizzare l'auto privata.					
Non mi interessa utilizzare mezzi di trasporto sostenibile nelle prossime due settimane.					
Sarebbe facile per me utilizzare il trasporto sostenibile.					
Sono certo di poter utilizzare il trasporto sostenibile nel corso della prossima settimana.					
Utilizzare il trasporto sostenibile è per me impossibile.					
Per me utilizzare i mezzi di trasporto sostenibile anziché l'auto privata è UTILE.					
Per me utilizzare i mezzi di trasporto sostenibile anziché l'auto privata è (o sarebbe) PIACEVOLE.					
Per me utilizzare i mezzi di trasporto sostenibile anziché l'auto privata è GIUSTO.					
Sarebbe facile per me utilizzare il trasporto sostenibile.					
Sono certo di poter utilizzare il trasporto sostenibile nel corso della prossima settimana.					
Utilizzare il trasporto sostenibile è per me impossibile.					

INFORMAZIONI PERSONALI
Indichi la sua età

Genere
- Maschio
- Femmina

Qual è la sua occupazione attuale?
- Studente universitario
- Specializzando/Dottorando/Assegnista/Borsista
- Lavoratore dipendente
- Lavoratore autonomo
- Casalinga
- Pensionato
- Disoccupato
- Altro

Qual è il suo titolo di studio?
- Licenza elementare
- Diploma di scuola media inferiore
- Diploma di scuola media superiore
- Specializzazione professionale
- Titolo Universitario (1-2 livello)
- Titolo Post-laurea (dottorato, specializzazione, etc.)

Da quante persone è composto il suo nucleo familiare, incluso lei?

Ha figli che vivono nel suo nucleo familiare?
- No
- Sì (specifichi il numero)

Indichi il numero di componenti del suo nucleo familiare di età inferiore ai 10 anni

Ha la patente?
- Sì
- No

Ha una bicicletta di sua proprietà o a sua disposizione (ad es. bici di famiglia) per i suoi spostamenti?
- Sì
- No

Ha un'auto di sua proprietà o a sua disposizione (ad es. auto di famiglia) per i suoi spostamenti?
- Sì
- No

In totale, di quante auto disponete in famiglia?
- 0
- 1
- 2
- 3
- 4
- 5 o più

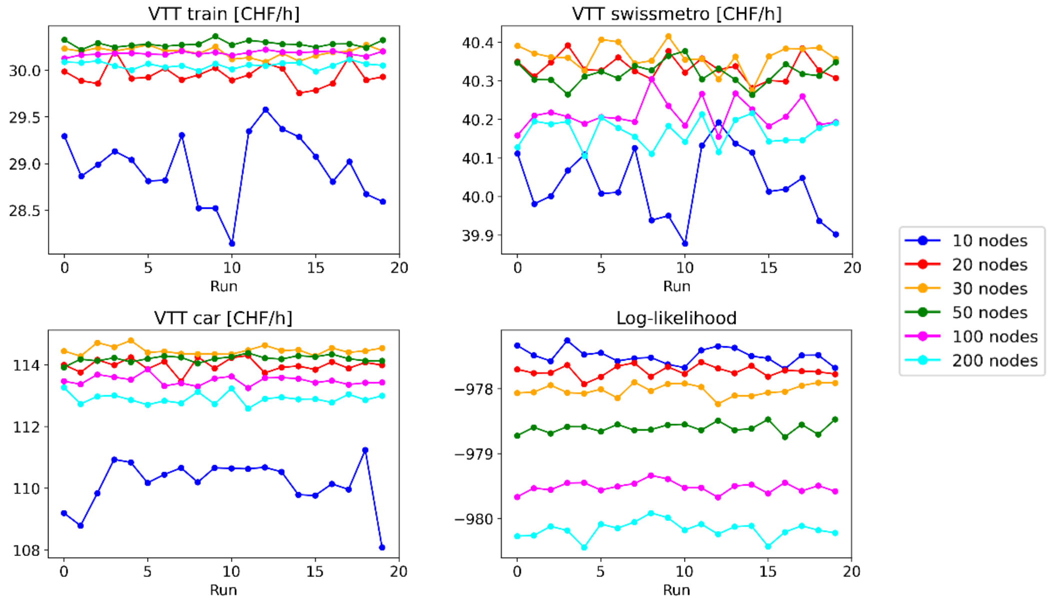
Potrebbe indicare in quale fascia di reddito mensile netto INDIVIDUALE si riconosce?
- Non percepisco reddito
- Inferiore a 500 €
- Tra 500 e 1000 €
- Tra 1000 e 1500 €
- Tra 1500 e 2000 €
- Tra 2000 e 3000 €
- Maggiore di 3000 €

Pur non ricevendo un reddito, indichi qual è il budget mensile (€) di cui dispone per sé:

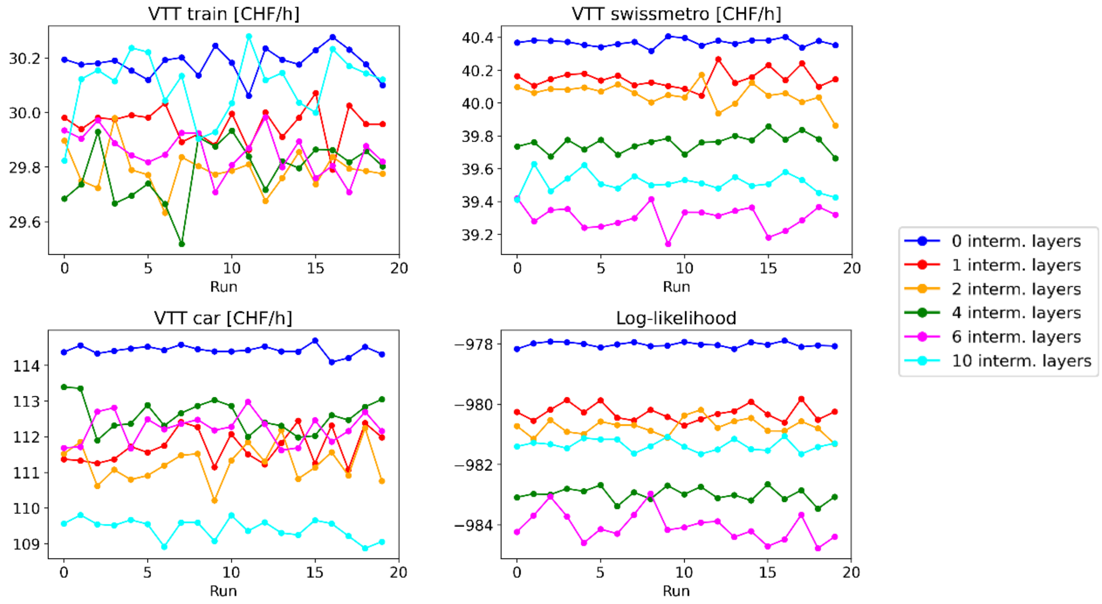
Quale sistema operativo nel suo smartphone possiede?
- Android (Samsung, HTC, Huawei, Sony, ...)
- iOS (iPhone)
- Windows Phone
- Altro

Come è venuto a conoscenza del presente questionario?
- Cartolina dell'indagine
- Poster dell'indagine
- Sito web e pagina Facebook dell'indagine
- Invito via e-mail
- Siti web istituzionali
- Social media (Quotidiani, TG, etc.)
- Passaparola
- Altro (specifichi nel riquadro sottostante):

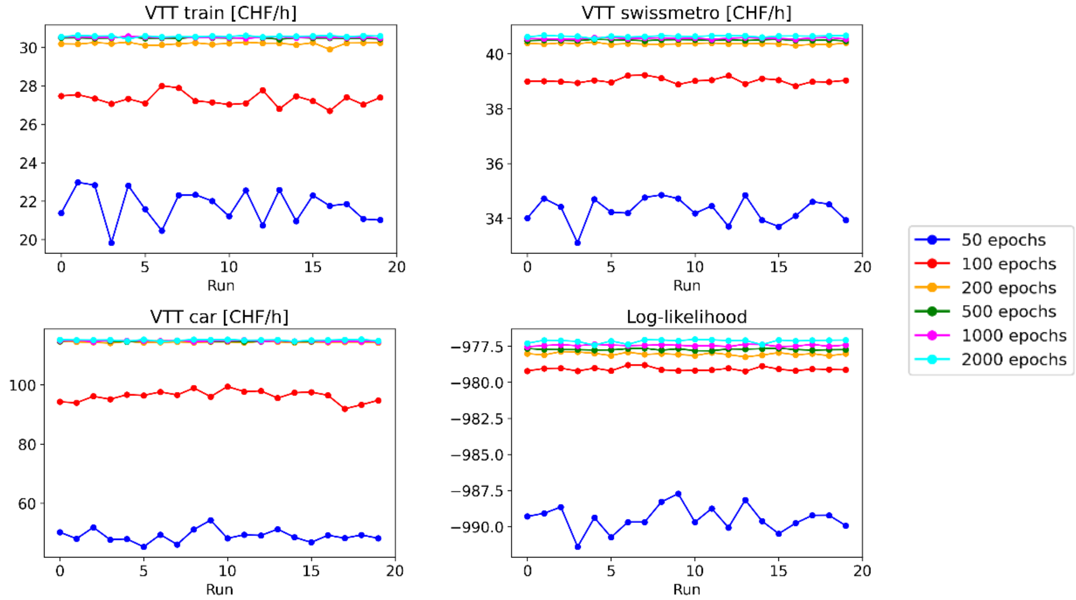
Appendix B



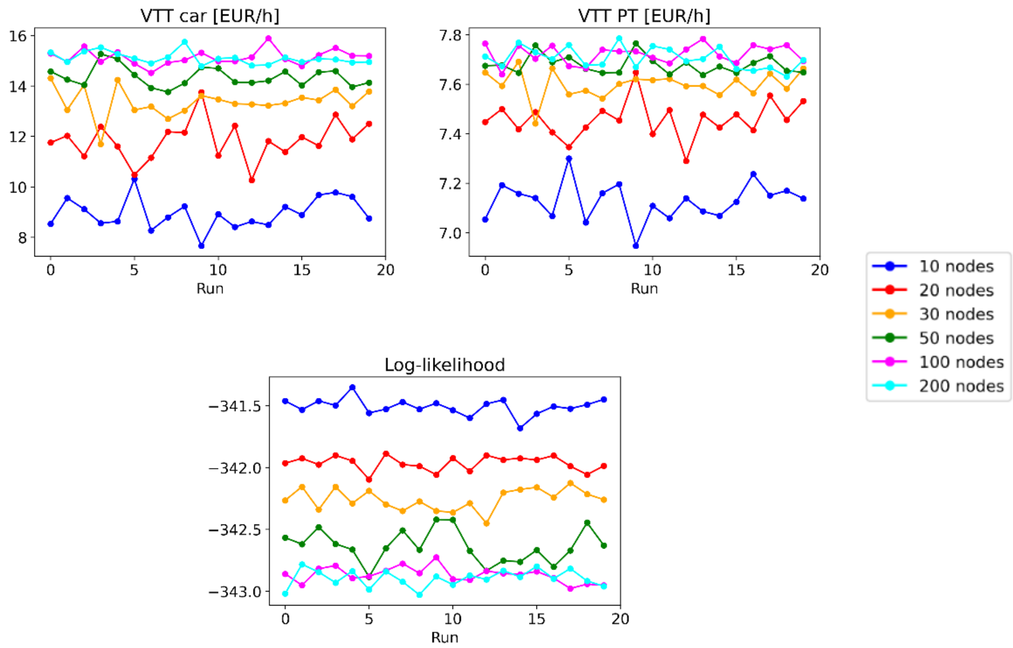
Swissmetro – nodes



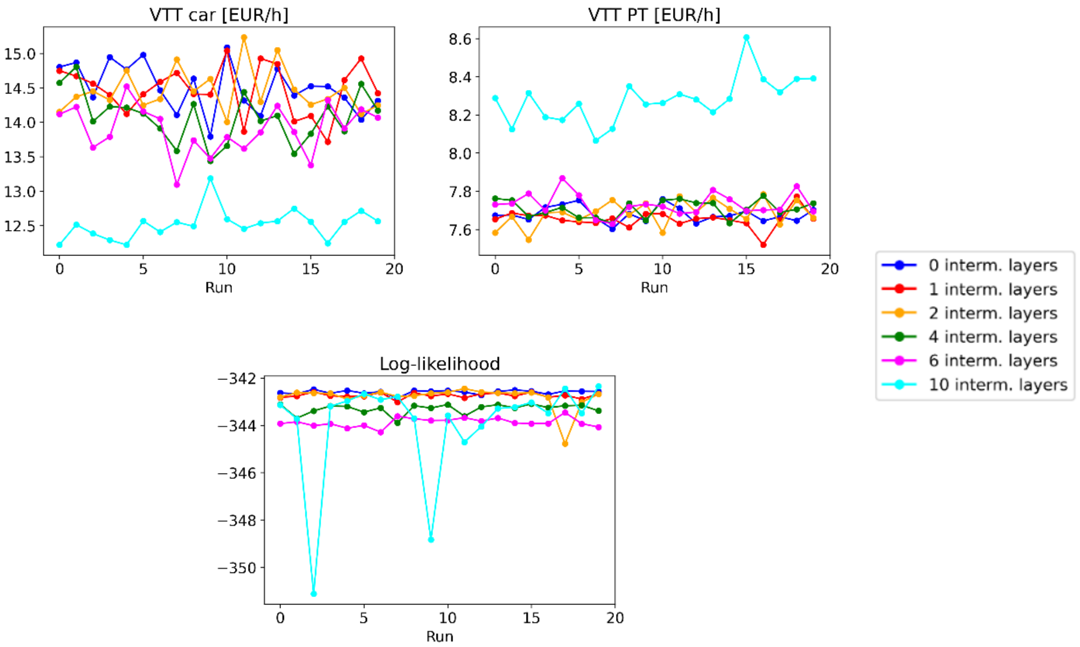
Swissmetro – layers



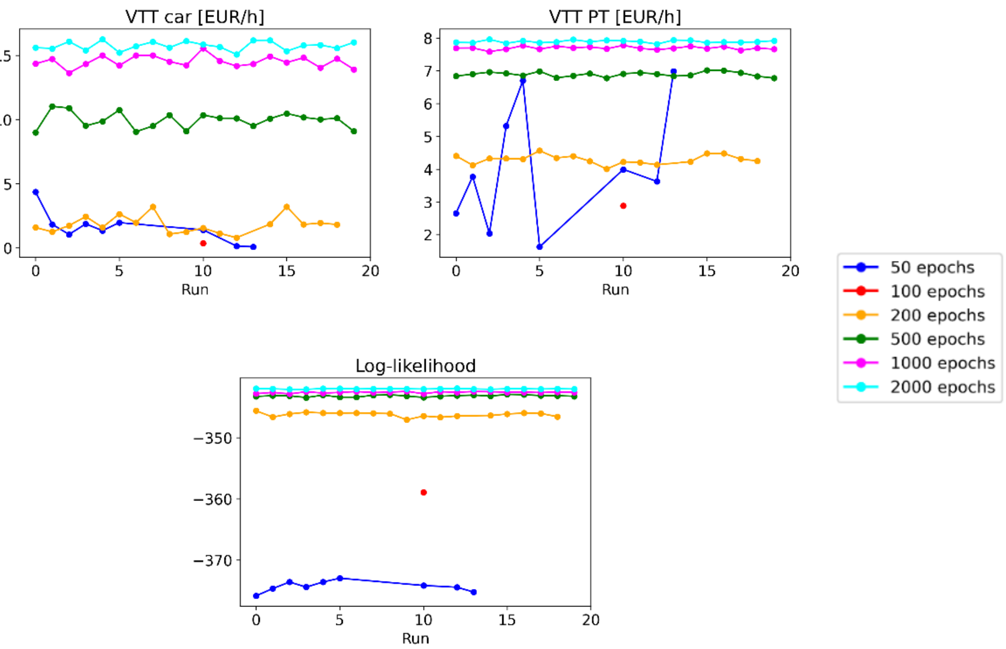
Swissmetro – epochs



Svolta Cagliari - nodes

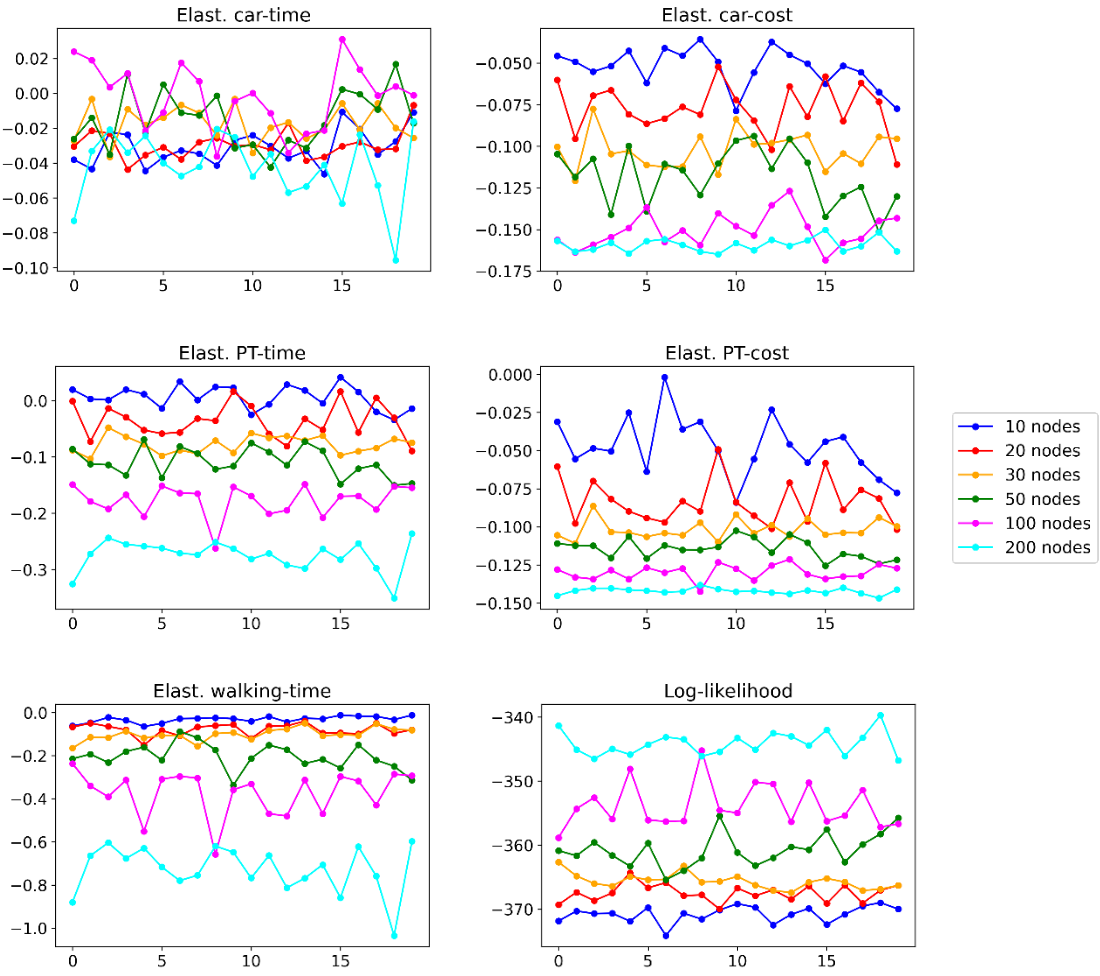


Svolta Cagliari - layers

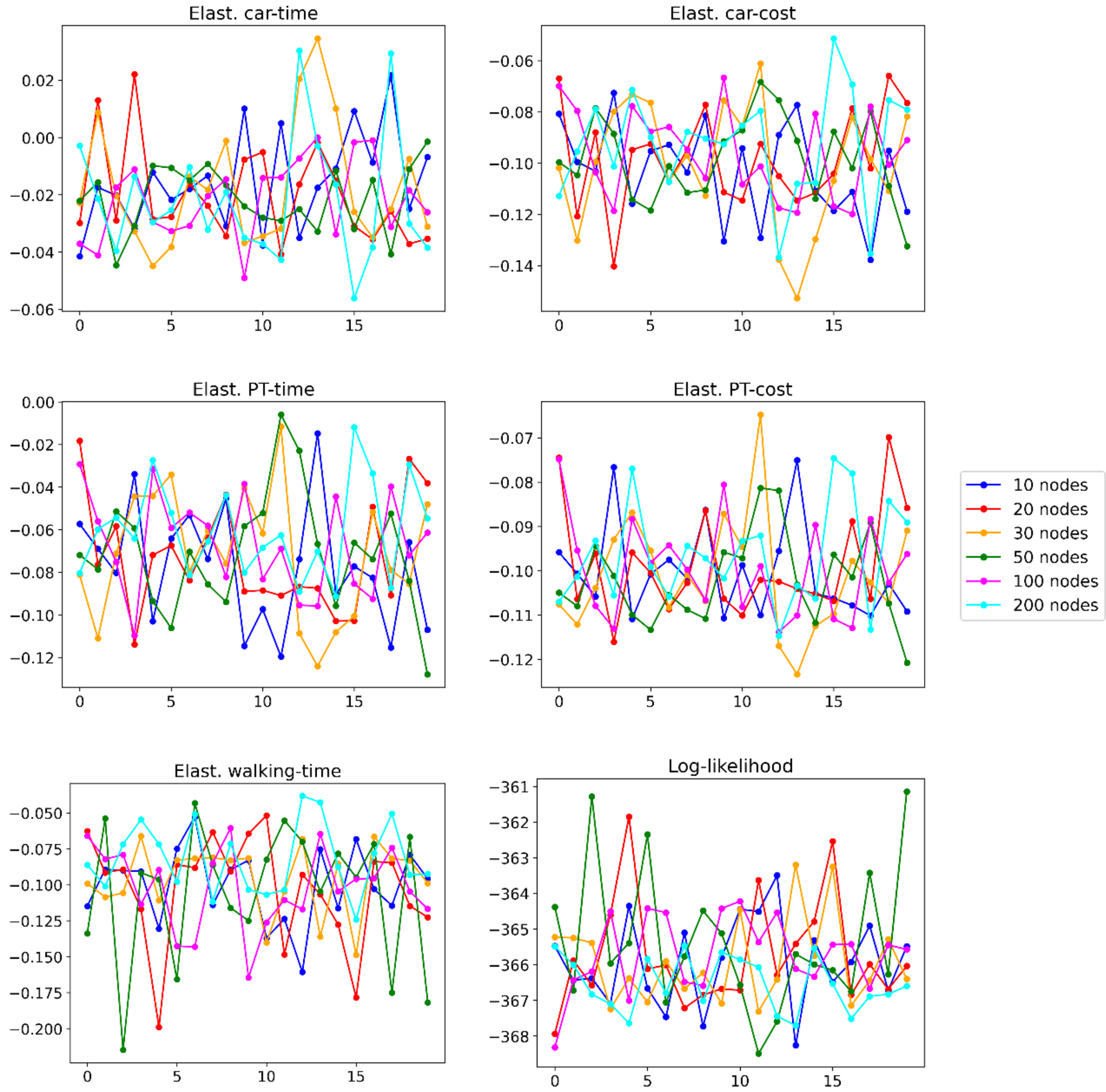


Svolta Cagliari - epochs

Appendix C



NN for ICLV – nodes (choice model)



NN for ICLV – nodes (latent model)

NN for ICLV – epochs