Modelling travel behaviour: A choice modelling perspective

Silvia Varotto [*, ‡]   Rico Krueger [†, ‡]   Michel Bierlaire [*]

15 March 2022

Report TRANSP-OR 220315
Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne
transp-or.epfl.ch

[*]École Polytechnique Fédérale de Lausanne (EPFL), School of Architecture, Civil and Environmental Engineering (ENAC), Transport and Mobility Laboratory, Switzerland, {silvia.varotto,michel.bierlaire}@epfl.ch

[†]Technical University of Denmark (DTU), Department of Technology, Management and Economics, Denmark, {rickr}@dtu.dk

[‡]Shared first authorship.

Abstract

Choice models have been applied to explain and predict the transportation choices of individuals for half a century. The advent of big data brings about new opportunities and poses new challenges for forecasting. This chapter discusses the major methodological contributions and the most recent developments in the field of choice modelling in transportation. Advanced choice models have been proposed to accommodate unrestricted substitution patterns between alternatives, unobserved taste variations, serial correlation between repeated observations, and latent constructs as attitudes and perceptions. In recent years, data-driven methods have gained traction to improve the prediction accuracy and to assist the analyst in the model specification. Choice models have also been incorporated into optimization problems to account for the interactions between the choices of individuals and the planning decisions under evaluation. To estimate these advanced models, fast and computationally efficient methods are required.

Keywords: transportation, advanced choice models, data-driven methods, choice-based optimization, estimation methods.

1 Introduction

Disaggregate behavioural models have been applied to travel demand analysis since 1970 (McFadden, 2001a,b). At that time, major changes in the transportation system that could not be reliably predicted using the aggregate models available were under evaluation at the US Department of Transportation. To address this challenge, Domencich and McFadden (1975) estimated a travel demand model using disaggregated utilities based on the *conditional logit model* developed by McFadden (1968, 1973). The data were based on the real choices of individuals who were surveyed on their travel behaviour via interviews. The forecasting abilities of these models were tested in practice to predict travel behaviour following the introduction of the new public transport system BART in San Francisco (McFadden et al., 1977). The forecasts resulted to be quite accurate compared to the predictions of the aggregate models previously used. Based on this evidence, the methods developed started to be adopted for transportation analysis worldwide.

In the last decades, the advent of new technologies for data collection in real time has brought about new opportunities and posed new challenges for forecasting. The location of individuals can be tracked in details using GPS data and be used as inputs for route choice models (Axhausen et al., 2004). The behaviour of road users can be observed in on-road studies with instrumented vehicles (Varotto et al., 2018, 2021), driving simulator experiments (Paschalidis et al., 2019) and virtual reality experiments (Bogacz et al., 2021; Nuñez Velasco et al., 2019). These data collection methods allow to investigate user responses to new technologies as automated vehicles, incorporate behavioural constructs and human factors, and analyse different choice situations over time for each individual. A major challenge is to develop mathematical models that can accommodate these data and forecast user behaviour in real time in Intelligent Transportation Systems.

This chapter discusses the major methodological contributions and the most recent developments in the field of choice modelling in transportation. Despite being widely used, the logit model has some major limitations because it assumes proportional substitution patterns, no unobserved taste heterogeneity, independence across repeated observations over time, and homoscedastic errors across alternatives. To address these limitations, advanced choice models have been developed. Some of these advanced models lack a closed form expression of the choice probabilities and thus need to be estimated using simulation methods. Choice models can be used to obtain aggregated forecasts at the population level and support decision-makers (e.g. the expected market shares of competing transport services).

In recent years, several studies have proposed data-driven models, estimation methods and validation techniques in transportation demand analysis (Hillel et al., 2021). To date, theory-based choice models are still preferred to data-driven approaches because the model parameters are directly interpretable, forecasting indicators can be derived from the model parameters, and forecasting in new settings is underpinned by the causality link provided by the theory (van Cranenburgh et al., 2021). However, data-driven methods offer certain advantages in terms of model specification, model selection and prediction accuracy. A promising research direction is to integrate these methods into choice models to improve current practices in transportation demand analysis.

Individual-level demand specified by discrete choice models can be aggregated over individuals to calculate demand functions at the market level and directly link the prices to the market shares (Koppelman, 1976; McFadden and Reid, 1975). The interaction of these demand functions with the supply determines the market equilibrium prices. To explicitly account for these interactions, discrete choice models describing the demand can be directly integrated into *choice-based optimization* problems. Currently, a promising research direction to improve the behavioural realism of these models is to integrate advanced discrete choice models into optimization problems.

The chapter is structured as follows. Section 2 describes the origins of RUM theory, the logit model, and the earliest methods for collecting data regarding the travel behaviour of individuals. Section 3 discusses advanced models explicitly capturing unrestricted substitution patterns between alternatives, unobserved taste heterogeneity, serial correlation between repeated observations over time, heteroscedastic errors across alternatives, and latent constructs as attitudes and perceptions. Section 4 presents maximum likelihood methods and Bayesian methods that can be used to estimate choice models. Section 5 describes how choice models can be aggregated to obtain forecasts at the population level as market shares and how indicators useful for policy analysis can be derived from choice models. Section 6 presents studies proposing data-driven methods as tools to search for the most suitable model specification and studies integrating data-driven models to achieve higher goodness-of-fit and prediction accuracy. Section 7 describes how choice models can be incorporated into optimization problems to account for the interactions between the choices of individuals and the planning decisions under evaluation. Section 8 discusses open challenges and directions for future research.

2 Foundations

2.1 Origins of RUM theory and logit model

The logit model has a mathematical formulation suitable for econometric applications and is based on the theories proposed in the field of mathematical psychology by Thurstone (1927) and on the choice axiom proposed by Luce (1959). Thurstone (1927) developed the *law of comparative judgement* to represent imperfect discrimination of individuals in choice tasks. He assumed that an individual n perceives an alternative j as

$$U_{jn} = f(\theta_{jn}) + \varepsilon_{jn}, \tag{1}$$

where $f(\theta_{jn})$ is a function of unknown parameters θ_{jn} and characteristics of the alternative j, and ε_{jn} is an independent normally distributed error term which is individual- and alternative-specific. Given two alternatives j and j' in the choice set, an individual will prefer alternative j over alternative j' if $(U_{jn} - U_{j'n}) > 0$. The probability that an individual n chooses alternative j is equal to

$$P(j|\theta_{jn}, \sigma_{j,j'}) = \Phi\left(\frac{f(\theta_{jn}) - f(\theta_{j'n})}{\sigma_{j,j'}}\right), \tag{2}$$

where $f(\theta_{jn}) - f(\theta_{j'n})$ is the mean and $\sigma_{j,j'}$ is the standard deviation. This form is now called binomial probit. Notably, the law of comparative judgement was generalized to stochastic utility maximization over multiple alternatives and defined as *random utility maximization* (RUM) model by Marschak (1960).

Luce (1959) postulated that, given $P(j|C_n)$ the probability that an individual n chooses alternative j in a set of mutually exclusive and exhaustive alternatives C_n, the ratio of choice probabilities for alternatives j and j' is constant for every choice set C_n which comprises the alternatives j and j' such that

$$\frac{P(j|C_n)}{P(j'|C_n)} = \frac{P_{j,j'}(j)}{P_{j,j'}(j')}. \tag{3}$$

The axiom was defined as *Independence from Irrelevant Alternatives* (IIA). Luce showed that, if this axiom holds, a positive "strict utility" V_{jn} can be associated with each alternative as follows

$$P(j|C_n) = \frac{V_{jn}}{\sum_{j' \in C_n} V_{j'n}}. \tag{4}$$

McFadden (1968, 1973) proposed a parametric exponential function for this strict utility

$$V_{jn} = \exp(X_{jn}^{\top}\beta), \tag{5}$$

where X_{jn} are the attributes of alternative j for individual n and β are parameters to be estimated. This formulation established, for the first time, a link between the RUM-concept and the specification of empirical travel demand models. The formulation was called conditional logit model because its ratio form was similar to the form of conditional probabilities and, in case of two alternatives, it corresponded to a logistic. This form is now called the logit model. The logit model was used by Domencich and McFadden (1975) to predict the mode of work commuting trips and the generation, destination and mode of shopping trips. The different choices were linked using inclusive values which were function of the choice probabilities at the lower level of the decision tree.

2.2 Data collection methods

The earliest applications in travel demand analysis in the 1970's were based on data of individuals which were surveyed on their travel behaviour via home or phone interviews (McFadden, 2001a). Respondents usually reported the trips executed, the transport mode, the travel time, the travel costs and other relevant trip characteristics. The attributes of all alternatives available to each individual in the choice task were usually calculated by the analysts based on the transportation network. This method provided *revealed preference* (RP) data from real choices of individuals.

During the last decades, transportation analysts have developed more cost-effective methods to select the sample in surveys based on the choices under evaluation and the observed behaviour of a specific group of individuals. These methods are called *choice-based sampling* methods. When analysing changes in the transportation system caused by the introduction of a new facility, the analyst can survey the individuals who are most likely to use that facility and compare them to a control-group of individuals surveyed at home.

A second important innovation is represented by the collection of *stated preference* (SP) data, a research direction which has been extensively investigated (Ben-Akiva et al., 2019). SP data can be collected using *conjoint analysis*, which involves the presentation of hypothetical choice tasks in an experimental design (Green and Srinivasan, 1978). The main advantages of this method are that it allows to specify the choice setting precisely, to study the introduction of new services, and to collect a large number of observations with moderate costs. The main limitation is that the level of realism might be limited. To forecast the impact of changes in the transportation system, SP data can be combined with RP data in a unified modelling framework (McFadden, 1986; Ben-Akiva and Morikawa, 1990; Morikawa et al., 1991; Hensher et al., 1998; Louviere et al., 1999).

3 Advanced models

3.1 Nested logit and MEV

The nested logit model accommodates flexible substitution patterns by grouping alternatives that share unobserved characteristics into nests. The red bus/blue bus paradox illustrates the unrealistic substitution patterns implied by the IIA property. Suppose there are two transportation mode choice alternatives with the same travel time, namely car and blue bus. A logit model that only considers travel time as explanatory variable predicts that the probability of choosing car is the same as the probability of choosing bus, i.e. $P(\text{car}) = P(\text{blue bus}) = 0.5$ such that $P(\text{car})/P(\text{blue bus}) = 1$. Now suppose that the public transport authority decides to procure red busses to complement the existing fleet of blue busses. One would expect that car still accounts for 50% of the mode share, while red bus and blue bus each account for 25% of the mode share with a combined share of 50%, i.e. $P(\text{car}) = 0.5$ and $P(\text{blue bus}) = P(\text{red bus}) = 0.25$ such that $P(\text{bus}) = P(\text{blue bus}) + P(\text{red bus}) = 0.5$. However, logit predicts that the three modes have an identical market share of $1/3$, i.e. $P(\text{car}) = P(\text{blue bus}) = P(\text{red bus}) = 1/3$ with $P(\text{bus}) = P(\text{blue bus}) + P(\text{red bus}) = 2/3$. This is because the IIA property implies proportional substitution, i.e. $P(\text{car})/P(\text{blue bus}) = 1$ must hold.
The nested logit groups alternatives into M nests indexed by $m = 1, \dots, M$. The alternatives within a nest are assumed to share unobserved characteristics such that the utility errors of the alternatives are additively separable into nest- and alternative-specific terms. The random utility of alternative $j \in C$ is then given by

$$U_{jn} = V_{jn}(X_{jn}, \beta) + \varepsilon_{m(j),n} + \varepsilon_{jn}, \tag{6}$$

where V_{jn} is the systematic utility of alternative j for individual n, which depends on explanatory variables X_{jn} and parameters β. $\varepsilon_{m(j),n}$ and ε_{jn} are the nest- and the alternative-specific error components, respectively. $m(j)$ is a mapping from alternatives to nests such that $m(j) = m$ if alternative j is in nest m. The total utility error of alternative j is $\tilde{\varepsilon}_{jn} = \varepsilon_{m(j),n} + \varepsilon_{jn}$. It is assumed to be extreme value type I distributed with location zero and scale μ, i.e. $\tilde{\varepsilon}_{jn} \sim EV_1(0, \mu)$. The alternative-specific error components are also extreme value type I distributed but with location zero and scale $\mu_{m(j)}$, i.e. $\varepsilon_{m(j),n} \sim EV_1(0, \mu_{m(j)})$. As a consequence, the variance-covariance matrix of the random utilities is block-diagonal. We have

$$\text{Cov}(U_{jn}, U_{j'n}) = \begin{cases} \frac{\pi^2}{6\mu^2} & \text{if } j = j', \\ \frac{\pi^2}{6\mu^2} - \frac{\pi^2}{6\mu_m^2} & \text{if } j \neq j', j \text{ and } j' \text{ are in the same nest } m, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

The nested logit belongs to the multivariate extreme value (MEV) family of discrete choice models (McFadden, 1978). Under the MEV assumption, the probability of choosing alternative $j \in C$ conditional on explanatory variables X_{jn} and parameters $\theta = \{\beta, \mu\}$ is given by

$$P(j|X_n;\theta) = \frac{\Lambda_{jn}(X_n,\theta)}{\sum_{j' \in C} \Lambda_{j'n}(X_n,\theta)}, \tag{8}$$

where

$$\Lambda_{jn}(X_n,\theta) = e^{V_{jn}(X_{jn},\beta)+\ln G_{jn}(\psi_{1n},\ldots,\psi_{Jn};\mu)} \tag{9}$$

with

$$\psi_{jn} = e^{V_{jn}(X_{jn},\beta)} \tag{10}$$

and

$$G_{jn}(\psi_{1n},\ldots,\psi_{Jn};\mu) = \frac{\partial G}{\partial \psi_{jn}}(\psi_{1n},\ldots,\psi_{Jn};\mu). \tag{11}$$

Here, V_{jn} is the systematic utility component that depends on explanatory variables X_{jn} and parameter β. $G(\psi_{1n},\ldots,\psi_{Jn};\mu)$ is called a MEV generating function with parameter μ. For nested logit, we have

$$G(\psi_{1n},\ldots,\psi_{Jn};\mu) = \sum_{m=1}^{M} \left(\sum_{j=1}^{J_m} \psi_{jn} \right)^{\frac{\mu}{\mu_m}} \tag{12}$$

with $\mu = \{\mu, \mu_1, \ldots, \mu_M\}$. For nested logit to be consistent with RUM, we need to have $\mu, \mu_1, \ldots, \mu_M \geq 0$ and $\frac{\mu}{\mu_m} \leq 1$.
Another member of the MEV family is the cross-nested logit (CNL) model (Bierlaire, 2006) which allows for fuzziness in the nest membership of an alternative. CNL has the following generating function:

$$G(\psi_{1n},\ldots,\psi_{Jn};\mu,\alpha) = \sum_{m=1}^{M} \left(\sum_{j=1}^{J_m} \left(\alpha_{jm}^{1/\mu} \psi_{jn} \right)^{\mu_m} \right)^{\frac{\mu}{\mu_m}} \tag{13}$$

with $\mu = \{\mu, \mu_1, \ldots, \mu_M\}$, whereby $\mu, \mu_1, \ldots, \mu_M \geq 0$, $\frac{\mu}{\mu_m} \leq 1$ and $\alpha_{jm} \in [0,1]$. Here, α_{jm} captures the degree of membership of alternative j in nest m.
MEV models have been adopted in the analysis of various travel-related behaviours. For example, Forinash and Koppelman (1993) employed a nested logit model to capture similarities between transport modes in an analysis of business travellers' intercity mode choice behaviour. Furthermore, Cervero and Duncan (2008) used a nested logit model to jointly estimate the probabilities that someone lives near a rail stop and also commutes by rail. In a joint analysis of work and home location and commute mode choice, Abraham and Hunt (1997) used a nested logit model to capture similarities between transport modes. In an analysis of parking choice behaviour, Hunt and Teply (1993) devised a nested logit model representing similarities among on-street and off-street parking alternatives. Vovsha (1997) used a cross-nested logit model to capture similarities between pure and combined modes in an analysis of urban mode choice behaviour. Finally, in an analysis of pedestrian walking behaviour, Antonini et al. (2006) devised a cross-nested logit model to jointly explain speed and direction choices.

3.2 Mixtures of logit

Mixtures of logit models accommodate unobserved taste heterogeneity, unrestricted substitution patterns and correlation in unobserved factors over time. Due to this flexibility, mixtures of logit models (also called "mixed logit") have become workhorse methods in travel demand analysis. The ability to accommodate unobserved taste heterogeneity significantly enhances the flexibility of choice models. In many empirical applications, it is reasonable to assume that individuals exhibit varying sensitivities to attributes. For example, it is often unrealistic to assume that everyone is equally sensitive to travel time. Sensitivities to attributes may vary as a function of observable characteristics of the individual or of the choice situation. However, an analyst is unlikely to have access to all observable factors that influence taste heterogeneity. Therefore, taste heterogeneity is often unobservable from the analyst's point-of-view. Likewise, the ability to accommodate unrestricted substitution patterns significantly enhances the flexibility of choice models to accurately characterise demand. Finally, in many empirical applications, analysts are able to observe repeated choices from individuals. For example, modern data collection methods allow researchers to track individuals' transportation choices over time and the same unobserved factors may influence the individual's decisions in each of the repeated choice situations.

Mixtures of logit extend standard logit models hierarchically by allowing utility parameters to vary randomly across observational units. In mixtures of logit, the probability of choosing alternative $j \in C$ is given by

$$P(j|X_n, \theta) = \int_\beta \left(\frac{e^{V_{jn}(X_{jn}, \beta)}}{\sum_{j' \in C} e^{V_{j'n}(X_{j'n}, \beta)}} \right) f(\beta|\theta) d\beta. \tag{14}$$

Here, $\frac{e^{V_{jn}(X_{jn}, \beta)}}{\sum_{j' \in C} e^{V_{j'n}(X_{j'n}, \beta)}}$ is the standard logit choice probability evaluated at β. $f(\beta|\theta)$ is the density of β with parameter θ. f is referred to as the mixing distribution. It may be parametric, semi-parametric or non-parametric (Krueger et al., 2020; Vij and Krueger, 2017). The normal distribution is the most common mixing distribution. It has two parameters, a mean vector and a variance-covariance matrix. If the mixing distribution is discrete such that the support of β is finite, the choice probability (14) becomes

$$P(j|X_n, \beta, \pi) = \sum_{s=1}^{S} \pi_s \frac{e^{V_{jn}(X_{jn}, \beta_s)}}{\sum_{j' \in C} e^{V_{j'n}(X_{j'n}, \beta_s)}} \quad \text{with} \sum_{s=1}^{S} \pi_s = 1, \tag{15}$$

where β_s is the parameter that is associated with components $s = 1, \ldots, S$, and π_s represents the share of component s. It is possible to include a component membership model in which the probability that an individual n belongs to component s depends on individual-specific attributes Z_n. The component membership model can be formulated using logit such that

$$\pi_{sn} = P(s|Z_n, \gamma) = \frac{e^{Z_n^\top \gamma_s}}{1 + \sum_{s'=2}^{S} e^{Z_n^\top \gamma_{s'}}}. \tag{16}$$

Here, $\gamma = \{\gamma_2, \ldots, \gamma_S\}$ are unknown parameters. For identification, the first component is set as a reference component.

Mixtures of logit also allow for a flexible specification of error components to induce correlation across random utilities. Often, the error components are assumed to be normally distributed with mean zero and an unknown standard deviation (e.g. Walker et al., 2007). When there are B error components indexed by $b = 1, \ldots, B$, the random utility writes as

$$U_{jn} = \tilde{V}(X_{jn}, \beta) + \sum_{b=1}^{B} d_{jb}\sigma_b\xi_{bn} + \varepsilon_{jn}, \tag{17}$$

where d_{jb} is one if error component b is associated with alternative j and zero otherwise. Error component b induces correlation across all alternative $j \in C$ for which $d_{jb} = 1$. σ_b is the scale of error component b, and ξ_{bn} is a standard normal random variable. $\tilde{V}(X_{jn}, \beta)$ is a scalar function that depends on explanatory variables X_{jn} and parameters β. However, the model identification in error components specifications of mixture of logit is not simple and must be carefully considered in empirical applications (Walker et al., 2007).

Furthermore, mixtures of logit accommodate correlation in unobserved factors over time. Suppose that individual n is observed to make a sequence of T choices indexed by $t = 1, \ldots, T$. If the utility parameters are assumed to vary across individuals but not across the choice situations faced by one individual, the probability of observing individual n's sequence of choices $\{y_{1n}, \ldots, y_{Tn}\}$ is

$$P(y_{1n}, \ldots, y_{Tn}|X_n, \theta) = \int_\beta \left(\prod_{t=1}^{T} \frac{e^{V_{tjn}(X_{tjn}, \beta)}}{\sum_{j' \in C} e^{V_{tj'n}(X_{tj'n}, \beta)}} \right) f(\beta|\theta)d\beta \tag{18}$$

with $y_{tn} \in C \; \forall \; t \in \{1, \ldots, T\}$.

Mixtures of logit are now workhorse models in travel behaviour analysis. A non-exhaustive list of mixtures of logit applications includes the analysis of preferences for travel time savings (Hess et al., 2006), travel time and reliability (Small et al., 2005), congestion pricing (Bhat and Castelar, 2002), Mobility-as-a-Service (Caiati et al., 2020) and autonomous on-demand mobility (Krueger et al., 2016) as well as the study of mode choice behaviour (Bhat, 1997, 2000), trip timing decisions (Börjesson, 2008), household location decisions (Walker and Li, 2007) and driver choices to deactivate automation (Varotto et al., 2017a).

3.3 Integrated choice and latent variable models

In many empirical applications, including modelling the demand of modern mobility systems, socio-psychological elements such as attitudes, beliefs and perceptions can largely influence the decision-making process of individuals. For instance, sensitivity to travel time in sustainable transport modes may vary across individuals based on their attitude towards the environment. These constructs are not directly observable but can be indirectly measured using psychometric indicators. To capture the impact of these constructs and enhance the explanatory power of choice models, the RUM framework introduced in Section 2 should be extended.

The integrated choice and latent variable (ICLV) model (Ben-Akiva et al., 2002; Walker, 2001) allows for the inclusion of latent constructs as explanatory variables in choice models. ICLV models consist of two components, a choice model and a latent variable model.

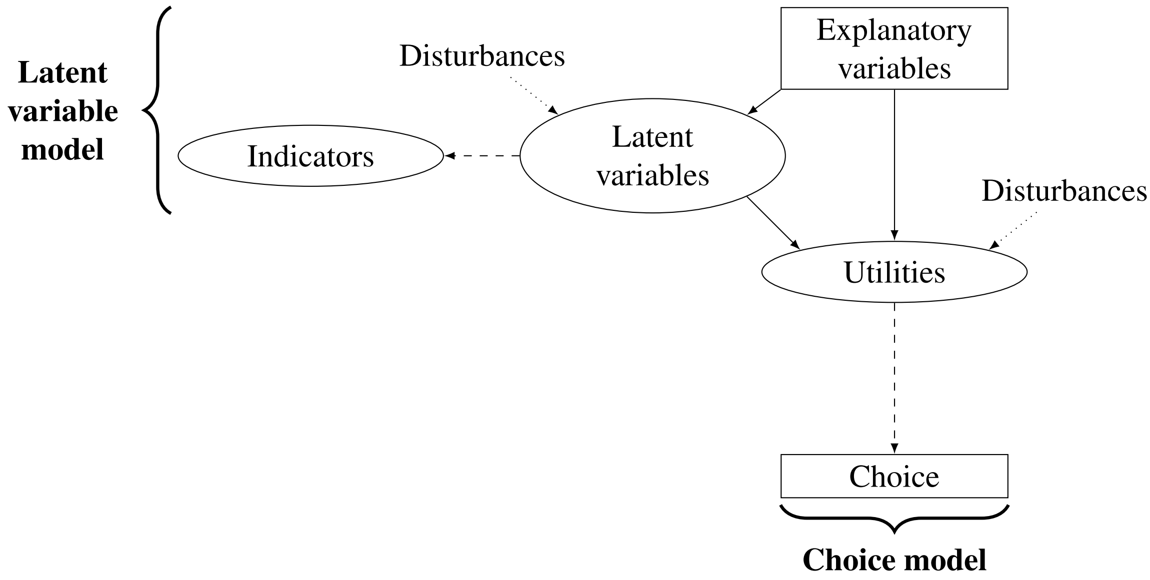

Figure 1: Integrated choice and latent variable model

Figure 1 visualises the structure of an ICLV model. In the choice model, the utility of alternative j is given by

$$U_{jn} = V_{jn}(X_{jn}, \phi_n; \beta), \tag{19}$$

where the systematic utility V_{jn} depends on observed attributes X_{jn}, latent variables ϕ_n and parameters β. The latent variable model consists of a structural component and a measurement component. We suppose that the model contains K latent variables indexed by $k = 1, \dots, K$ and L measurement indicators indexed by $l = 1, \dots, L$. The structural component of the latent variable model is given by

$$\phi_{kn} = W_{kn}(Z_{kn}, \gamma_k) + \eta_{kn}. \tag{20}$$

Here, W_{kn} is a function which depends on individual-specific characteristics Z_{kn} and parameters γ_k. η_{kn} is a random disturbance. The measurement component of the latent variable model is given by

$$m_{ln} = \delta_l + \phi_n^\top \lambda_l + \xi_{ln}, \tag{21}$$

where δ_l is a constant, λ_l is vector of factor loadings, and ξ_{ln} is an error term. m_{ln} denotes the measurement indicator. Depending on the nature of the indicator, the distribution of m_{ln} can be discrete or continuous. Then, the joint probability of the observed choice and the indicators is

$$P(y, m|X, \beta, \gamma, \delta, \lambda) = \int_\phi P(y|X, \phi, \beta)P(m|\delta, \phi, \lambda)f(\phi|Z, \gamma)d\phi. \tag{22}$$

Identification in ICLV models is non-trivial (e.g. Daly et al., 2012) and should be carefully considered in empirical applications.
ICLV models have been applied in various ways to analyse the relationship between different latent variables and travel behaviour. For example, Atasoy et al. (2013) used an

ICLV model to examine the influence of attitudes towards public transportation and environmental problems on transport mode choice. Paulssen et al. (2014) developed a hierarchical ICLV model to investigate the relationship between values, attitudes and transport mode choice. Motoaki and Daziano (2015) employed an ICLV model to analyse the influence of cycling skills and experience on cycling route choice preferences. Li and Kamargianni (2020) adopted an ICLV model to investigate the influence of attitudes and perceptions on shared mobility usage. Furthermore, Varotto et al. (2017b) used an ICLV model to account for error in travel time reporting in the analysis of revealed preference mode choice data.

4 Estimation

4.1 Maximum likelihood estimation

Point estimates of the parameters of choice models with closed form choice probabilities such as logit and other MEV-based models can be obtained using maximum likelihood estimation.

Suppose that we analyse a sample of N individuals indexed by $n = 1, \dots, N$. Every individual in the sample is observed to choose an alternative y_n out of the set $\mathcal{C} = 1, \dots, J$. We further assume a parametric form for a discrete choice model generating the probability that individual n chooses alternative $j \in \mathcal{C}$:

$$P(j|X_n; \theta). \tag{23}$$

Then, the log-likelihood of the sample is

$$\mathcal{L}(\theta) = \sum_{n=1}^{N} \ln P(j|X_n; \theta), \tag{24}$$

and a point estimate $\hat{\theta}$ of θ is given by

$$\hat{\theta} = \arg\max_{\theta} \mathcal{L}(\theta). \tag{25}$$

This optimization problem can be solved using standard optimization routines such as quasi-Newton methods.

However, the choice probabilities of continuous mixtures of logit and ICLV models are not analytically tractable and thus need to be approximated. For example, in continuous mixtures of logit, the intractable choice probability (14) can be approximated using R simulation draws denoted by β_{rn} with $r = 1, \dots, R$ from the density $f(\beta_n|\theta)$:

$$P(j|X_n, \theta) \approx \frac{1}{R} \sum_{r=1}^{R} \left(\frac{e^{V_{jn}(X_n, \beta_{rn})}}{\sum_{j' \in C} e^{V_{j'n}(X_n, \beta_{rn})}} \right). \tag{26}$$

Then, the simulated log-likelihood function of mixtures of logit is

$$\mathcal{L}(\theta) = \sum_{n=1}^{N} \ln \left(\frac{1}{R} \sum_{r=1}^{R} \left(\frac{e^{V_{jn}(X_n, \beta_{rn})}}{\sum_{j' \in C} e^{V_{j'n}(X_n, \beta_{rn})}} \right) \right). \tag{27}$$

The variance of the approximation defined in (26) decreases when R increases. Since the computational budget is usually finite, it is desirable to decrease the simulation variance for fixed R. Given an identical number of draws, strategically generated simulation draws such as quasi-random sequences incur a lower simulation error than pseudo-random draws. The field of travel demand modelling has contributed significantly to the development and understanding of quasi-random simulation techniques (Bhat, 2003; Hess et al., 2006; Sivakumar et al., 2005). Popular methods include Halton draws (Bhat, 2003) and Modified Latin Hypercube Sampling (Hess et al., 2006).

4.2 Bayesian methods

Maximum simulated likelihood estimation presents a powerful framework for the estimation of continuous mixtures of logit. However, the computational cost of maximum simulated likelihood estimation becomes exceedingly large when models contain complex hierarchies of latent variables and parameters. Bayesian estimation is an alternative approach for the estimation of choice models which alleviates the issues of maximum simulated likelihood estimation.

A key difference between maximum (simulated) likelihood estimation and Bayesian estimation is that Bayesian estimation aims to infer the posterior distribution of unknown parameters, whereas maximum (simulated) likelihood estimation aims to find point estimates of unknown parameters. Bayesian estimation entails the specification of a full probability model defining the joint distribution of the observed data y and the unknown quantities (including parameters and latent variables) θ. This joint distribution has the form $P(y, \theta) = P(y|\theta)P(\theta)$. Here, $P(y|\theta)$ is the likelihood of y given θ. $P(\theta)$ is the prior distribution of θ, which captures the state of knowledge about the distribution of the unknown model quantities before observing the data y. The posterior distribution $P(\theta|y)$ which we wish to estimate captures the state of knowledge about the distribution of the unknown model quantities θ after observing the data y. By Bayes' rule, the posterior distribution of interest is given by

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{\int P(y|\theta)P(\theta)d\theta} \propto P(y, \theta). \tag{28}$$

The posterior distributions of MEV- and probit-based discrete choice models are not available in closed form. Hence, posterior inferences in these types of models are performed using approximate approaches such as Markov chain Monte Carlo (MCMC) and Variational Bayes (VB).

MCMC methods approximate an intractable posterior distribution through samples from a Markov chain such that the stationary distribution of these samples is the posterior distribution of interest. The Metropolis-Hastings algorithm is the most widely used method to construct a Markov chain. The algorithm consists of two steps. First, a new state is drawn from a proposal distribution conditionally on the current state of the Markov chain. Then, with some probability, the new state is either accepted and used in the next iteration, or the new state rejected and the old state is re-used in the next iteration. Gibbs sampling is a special case of the Metropolis-Hastings algorithm. In Gibbs sampling, the unknown quantities of interest are divided into blocks, and one block of parameters is updated conditionally on all other parameter blocks. Gibbs sampling is appealing when the

conditional posterior distributions correspond to known distributions from which direct sampling is easy.
In particular in applications to large datasets, MCMC methods succumb to several limitations, namely the need to create sufficient storage for the posterior draws, the lack of a well-defined convergence criterion and autocorrelation of the generated posterior draws (Bansal et al., 2020). Variational inference seeks to overcome the limitations of MCMC by formulating approximate Bayesian inference as an optimization problem which consists of minimising the probability distance between the target posterior distribution $P(\theta|y)$ and a parametric variational distribution $q(\theta|\boldsymbol{\nu})$ over the parameter $\boldsymbol{\nu}$ of the variational distribution.
Compared to maximum simulated likelihood methods, Bayesian estimation methods have found far fewer applications in travel behaviour analysis, likely due to a steeper learning curve. Applications of MCMC methods in the context of discrete choice and travel behaviour analysis include the estimation of mixtures of logit models with flexible mixing distributions for analysing preferences for mobility on-demand (Krueger et al., 2020), integrated choice and latent variable models for analysing preferences for environmentally-friendly vehicle technologies (Daziano and Bolduc, 2013) and mixtures of logit models with unobserved inter- and intra-individual heterogeneity for analysing mode choice (Krueger et al., 2021). Also variational inference methods have been employed in the context of discrete choice and travel behaviour analysis. Bansal et al. (2020) compared different implementations of variational inference for mixtures of logit and apply the estimators to analyse preferences for alternative fuel vehicles. In an analysis of stated preferences for different transport modes, Rodrigues et al. (2020) relied on variational inference methods to estimate a logit model which automatically determines the most relevant predictors out of a large set of explanatory variables. Furthermore, Wong and Farooq (2020) employed variational inference methods to estimate discrete-continuous models of travel behaviour using large datasets.

5 Model application and forecasting

5.1 Travel demand prediction

In trip-based models, individual trips are used as unit of analysis and travel demand is usually modelled in four consecutive steps. The first step is trip generation which involves the calculation of the number of trips generated and attracted by each zone. The second step is the trip distribution that defines the number of trips from each zone of origin to each zone of destination (OD pair). The third step is the mode choice that splits the trips between OD pairs to the different modes of transport (e.g., passenger vehicles and public transport). The fourth step is the assignment that assigns the trips for each mode to the corresponding network (e.g., road network or public transport network). Each trip is assumed to be independent from the other trips. The forecasting abilities of RUM-models were tested in practice to predict travel behaviour following the introduction of the new public transport system BART in San Francisco (McFadden et al., 1977). The forecasts resulted to be quite accurate compared to the predictions of the aggregate models previously used. Based on this evidence, the methods developed started to be adopted for transportation analysis worldwide. Notably, these disaggregated models were aggregated

at the zonal level before the assignment and they did not explicitly capture the departure time choice.
A major limitation of trip-based models is to consider individual trips as the unit of analysis. In practice, the location, the travel mode, and the departure time of the different trips executed by the same individual are influenced by some common characteristics and the different trips cannot be analysed separately (Bhat and Koppelman, 1999). In addition, individual travelers in the same household might influence each other. These limitations have been addressed by adopting activity-based models (Axhausen and Gärling, 1992). In these models, travel demand is considered derived from the desire or the need to pursue certain activities. These models allow to examine how individuals change their participation in activities in response to changes in the travel conditions during a certain period of time which is used as unit for analysis (e.g., one working day). These models can incorporate high levels of spatial and temporal resolution, and can explicitly accommodate both individual activity behaviour and links between different individuals in a household. The models are usually implemented into a micro-simulation framework, in which the choices of individuals and households are assessed. These models can be easily extended to evaluate the impact of new policies that cannot be represented in trip-based models (e.g., a novel pricing alternative when entering a specific area for the first time and no additional pricing when returning).

5.2 Indicators for cost-benefit analysis

To measure the benefits associated with potential improvements in the transportation system and assess whether these benefits compensate the costs, researchers can calculate the change in *consumer surplus*. In travel demand analysis, the consumer surplus is defined as the difference between what travellers are willing to pay for an improvement in the system and what they pay in practice. The foundations for the calculation of the consumer surplus based on RUM models were provided by Williams (1977) and Small and Rosen (1981). The consumer surplus can be calculated as the area under the demand function (i.e., the choice probability) and above the market price (i.e., the utility, as multiple variables explain the behaviour). Researchers can be interested in calculating the change in consumer surplus determined by a change in the utilities from V^1 to V^2 and/or a change in the choice set from C^1 to C^2. For a logit model, the change in consumer surplus is given by

$$\frac{1}{\mu} \ln \sum_{j \in C^2} e^{\mu V^2} - \frac{1}{\mu} \ln \sum_{j \in C^1} e^{\mu V^1}. \tag{29}$$

For MEV models in general, the consumer surplus is given by

$$\ln G\left(e^{V^2}\right) - \ln G\left(e^{V^1}\right), \tag{30}$$

where G is the MEV generating function. The generating function for nested logit models is given in (12). In these equations, the utility can be transformed into monetary values dividing by the cost parameter.
Another useful indicator used in cost-benefit analysis is the *willingness-to-pay* (WTP) for improvements in the transportation system. The WTP is defined as the net income decrease that, when a certain variable is modified, allows to maintain the same expected

utility. In travel demand analysis, a typical WTP measure is the *value of time* (VOT) which is defined as the cost that travellers are willing to pay to save travel time. Using this measure, it is possible to compare the actual costs of an improvement in the transportation system to the expected benefits for the travellers, expressed in monetary terms. Williams (1977) and Daly and Zachary (1979) provided the foundations for the calculation of the WTP when the demand behaviour is represented by a RUM model. In this case, the VOT_{jn} for alternative j and individual n can be calculated using conventional measures of consumer surplus as follows

$$\mathrm{VOT}_{jn} = \frac{\partial V_{jn}/\partial tt_{jn}}{\partial V_{jn}/\partial c_{jn}}, \tag{31}$$

where V_{jn} is the systematic utility, tt_{jn} is the travel time, c_{jn} is the cost. When the variables travel time and cost are linearly included into the utility function, the VOT is constant and is given by

$$\mathrm{VOT} = \frac{\beta_{tt}}{\beta_c}, \tag{32}$$

where β_{tt} and β_c are the parameters associated with the travel time and the cost respectively.

6 Integration with data-driven methods

6.1 Automated selection of model structure and specification

In discrete choice models, the model structure and the explanatory variables are chosen by the analyst based on existing theories and knowledge on the data generation process. The final model structure and specification are the outcome of an extensive selection process based on formal testing between alternative structures and specifications. In contrast, data-driven methods allow to explore several model structures and specifications without relying on any *a priori* assumptions on the problem under investigation. These approaches are useful to assist the analyst, saving time during the model selection phase and exploring specifications that would not have been considered. In the literature, two relevant approaches are available: i) using data-driven methods to inform choice models and ii) translating the model specification task into an optimization problem.

The first approach consists in identifying key variables based on data-driven methods and then using the outputs of the exploration to inform the development of choice models. This approach was originally proposed in the field of marketing (Bentz and Merunka, 2000; Hruschka et al., 2002), where a neural network was used as a diagnostic and specification tool for the development of a choice model. In the field of transportation, relevant variables were identified using a gradient-boosting decision-tree (Hillel et al., 2019) and using the concept of automatic relevance determination in a Bayesian framework (Rodrigues et al., 2020).

The second approach consists in translating the model specification task into an optimization problem which can be solved based on search algorithms. Several studies have proposed different variable selection methods, designed to operate with the original variables and based on metaheuristics (for a review, we refer to Ortelli et al. (2021)). However, only a few studies have proposed flexible methods that can accommodate more complex model specifications. In the field of transportation, Paz et al. (2019) showed that metaheuristics as simulated annealing can be adapted to select both the variables and the parameter distributions in mixtures of logit model. A similar approach was proposed by Ortelli et al. (2021) who defined the model specification task as a multi-objective combinatorial optimization problem based on a set of information defined by the analyst. They implemented a variant of the variable neighbourhood search algorithm which explores several candidate models simultaneously. This method can accommodate any form of variable transformation and different model structures.

6.2 Goodness of fit and prediction accuracy

Several studies have shown that data-driven methods offer higher prediction accuracy than choice models (for an extensive comparison, we refer to Wang et al. (2021)). However, limited efforts have been dedicated to link these models with economic theories and extract interpretable results, which are fundamental requirements for making planning and policy decisions in transportation. Brathwaite et al. (2017) developed a microeconomic framework for the interpretation of bayesian model trees and integrated these models into choice models to capture semi-compensatory decision protocols. Inspired by methods in computer vision, Alwosheel et al. (2019) proposed synthesising prototypical examples to investigate the realism of the relationships learned by a neural network. Zhao et al. (2020) calculated marginal effects and arc elasticities using a modified approach to address the limitations of a random forest model and obtain more realistic outputs. In these studies, the machine learning models showed higher predictive power and a higher or lower level of behavioral realism than the choice models.

To achieve interpretable results and improve the prediction accuracy, some studies have proposed to combine choice models and data-driven methods in a unified framework. Sifringer et al. (2020) developed a logit model in which the utility function is manually specified by the analyst and the alternative-specific constants are modelled as functions of the variables not included in the utility function using a neural network. This approach was extended by Han et al. (2020), who allowed all parameters to be specified as a function of the characteristics of the individual using neural networks. Inspired by natural language processing approaches, Pereira (2021) proposed a method for encoding categorical variables using neural networks which learn embeddings of these variables. Finally, Wong and Farooq (2021) integrated a deep neural network into a choice model and extracted behavioural indicators from the matrix parameters. In these studies, the models integrating neural networks showed higher predictive power than the original choice models and allowed to directly interpret part of the utility function.

7 Choice-based optimization

In *choice-based optimization* problems, supply decisions are jointly optimized with demand decisions. Supply and demand decision influence each other until a global optimum solution satisfying both has been found. Demand decisions are represented by choice models that contain the supplier decision variables. Logit models do have a closed form expression of the choice probabilities that facilitates the integration into large-scale optimization problems and guarantees the equilibrium existence and uniqueness. Liu et al. (2019) developed a unified framework to optimize and analyze the operation of Mobility-On-Demand systems, where passenger mode choice is predicted using a logit model and the level of service of a travel mode influences its demand. A Bayesian optimization approach was used to calculate the optimal demand parameters. The framework allows to evaluate the impact of Mobility-On-Demand on the transportation system and vice versa. Advanced choice models, which have a higher level of behaviour realism, do not have a closed form expression and result in non-convex optimization problems. With these models, there is not guarantee of equilibrium existence and analytical approaches cannot be used to search for it. Recently, Pacheco Paneque et al. (2021) developed a new approach to integrate advanced choice models into a mixed integer linear optimization problem by using simulation. The simulation approach provides a realistic representation of the interactions between the individual choices and the variables under evaluation across a large number of simulation-based replications. To date, however, such approach can only be applied to solve small-scale problems because the simulations have a high computational cost. Based on this simulation approach, Bortolomiol et al. (2021b) developed an algorithm to find an approximate equilibrium solution in a market in which the demand is modelled at the individual level. The method can explicitly capture observed and unobserved heterogeneity between individuals in the demand function, multiple offers by the suppliers and different pricing strategies. Building on these previous studies, Bortolomiol et al. (2021a) developed a framework to identify optimal transport policies to regularize oligopolistic transport markets in which the demand is modelled at the individual level using choice models. In these markets, the regulations can influence the decisions of all agents involved. Similar numerical approaches are needed when the optimization problem is noncovex. Chakraborty et al. (2021) developed a framework for optimal design of exclusive lanes for automated vehicles on freeways which incorporates the demand split among automated vehicles and manually driven vehicles using a logit model. Due to the binary variables representing the lane design, the problem was formulated as a noncovex mixed-integer nonlinear program. The results show that, when accounting for the demand of each mode, the optimal lane design is not trivial.

8 Conclusion

Models for explaining and predicting the transportation choices of individuals have become indispensable in travel behaviour analysis. In this chapter, we reviewed major methodological advances and recent developments in the field of choice modelling in transportation.

Modern mobility systems, emerging data sources and machine learning create new needs and opportunities for the development of innovative methods for the choice-based analysis of travel behaviour. In particular, choice-based travel demand models need to be adapted to explain and predict increasingly complex travel patterns. Furthermore, optimization-based decision support tools need to be developed to enhance the efficiency of modern mobility systems using insights into user preferences and decision making.

To achieve these strategic research objectives, methods at the intersection of estimation methods, data-driven approaches and choice-based optimization need to be advanced.

First, growing amounts of data create a need for fast and computationally-efficient estimation methods for advanced choice models. Second, methods for the assisted specification of discrete choice models should be extended to accommodate advanced models with latent variables and latent classes and to improve the efficiency of existing algorithms by defining restrictions to the search space using knowledge from literature. Third, more studies on the interpretability of machine learning models and their integration into choice models are needed in order to obtain realistic forecasting indicators for travel demand analysis. Finally, efficient mathematical decomposition techniques need to be developed to speed-up choice based optimization problems using advanced choice models and apply these models to large-scale problems.

References

Abraham, J. E. and Hunt, J. D. (1997). Specification and estimation of nested logit model of home, workplaces, and commuter mode choices by multiple-worker households. *Transportation Research Record*, 1606(1):17–24.

Alwosheel, A., van Cranenburgh, S., and Chorus, C. G. (2019). 'computer says no' is not enough: Using prototypical examples to diagnose artificial neural networks for discrete choice analysis. *Journal of Choice Modelling*, 33:100186.

Antonini, G., Bierlaire, M., and Weber, M. (2006). Discrete choice models of pedestrian walking behavior. *Transportation Research Part B: Methodological*, 40(8):667–687.

Atasoy, B., Glerum, A., and Bierlaire, M. (2013). Attitudes towards mode choice in switzerland. *disP-The Planning Review*, 49(2):101–117.

Axhausen, K., Schonfelder, S., Wolf, J., Oliveria, M., and Samaga, U. (2004). Eighty weeks of gps traces, approaches to enriching trip information. In *Transportation Research Board Annual Meeting*, pages 1870–06. Citeseer.

Axhausen, K. W. and Gärling, T. (1992). Activity-based approaches to travel analysis: conceptual frameworks, models, and research problems. *Transport Reviews*, 12(4):323–341.

Bansal, P., Krueger, R., Bierlaire, M., Daziano, R. A., and Rashidi, T. H. (2020). Bayesian estimation of mixed multinomial logit models: Advances and simulation-based evaluations. *Transportation Research Part B: Methodological*, 131:124–142.

Ben-Akiva, M., McFadden, D., Train, K., Walker, J., Bhat, C., Bierlaire, M., Bolduc, D., Boersch-Supan, A., Brownstone, D., Bunch, D. S., et al. (2002). Hybrid choice models: Progress and challenges. *Marketing Letters*, 13(3):163–175.

Ben-Akiva, M. and Morikawa, T. (1990). Estimation of switching models from revealed preferences and stated intentions. *Transportation Research Part A: General*, 24(6):485–495.

Ben-Akiva, M. E., McFadden, D., Train, K., et al. (2019). *Foundations of stated preference elicitation: Consumer behavior and choice-based conjoint analysis*. Now.

Bentz, Y. and Merunka, D. (2000). Neural networks and the multinomial logit for brand choice modelling: a hybrid approach. *Journal of Forecasting*, 19(3):177–200.

Bhat, C. R. (1997). An endogenous segmentation mode choice model with an application to intercity travel. *Transportation science*, 31(1):34–48.

Bhat, C. R. (2000). Incorporating observed and unobserved heterogeneity in urban work travel mode choice modeling. *Transportation science*, 34(2):228–238.

Bhat, C. R. (2003). Simulation estimation of mixed discrete choice models using randomized and scrambled halton sequences. *Transportation Research Part B: Methodological*, 37(9):837–855.

Bhat, C. R. and Castelar, S. (2002). A unified mixed logit framework for modeling revealed and stated preferences: formulation and application to congestion pricing analysis in the san francisco bay area. *Transportation Research Part B: Methodological*, 36(7):593–616.

Bhat, C. R. and Koppelman, F. S. (1999). *Activity-based modeling of travel demand*, chapter 3, pages 35–61. International Series in Operations Research & Management Science. Springer, Boston.

Bierlaire, M. (2006). A theoretical analysis of the cross-nested logit model. *Annals of operations research*, 144(1):287–300.

Bogacz, M., Hess, S., Calastri, C., Choudhury, C. F., Mushtaq, F., Awais, M., Nazemi, M., van Eggermond, M. A., and Erath, A. (2021). Modelling risk perception using a dynamic hybrid choice model and brain-imaging data: An application to virtual reality cycling. *Transportation Research Part C: Emerging Technologies*, 133:103435.

Börjesson, M. (2008). Joint rp–sp data in a mixed logit analysis of trip timing decisions. *Transportation Research Part E: Logistics and Transportation Review*, 44(6):1025–1038.

Bortolomiol, S., Lurkin, V., and Bierlaire, M. (2021a). Price-based regulation of oligopolistic markets under discrete choice models of demand. *Transportation*. Accepted on Jul 23, 2021.

Bortolomiol, S., Lurkin, V., and Bierlaire, M. (2021b). A simulation-based heuristic to find approximate equilibria with disaggregate demand models. *Transportation Science*. Accepted on Apr 16, 2021.

Brathwaite, T., Vij, A., and Walker, J. L. (2017). Machine learning meets microeconomics: The case of decision trees and discrete choice. *arXiv preprint arXiv:1711.04826*.

Caiati, V., Rasouli, S., and Timmermans, H. (2020). Bundling, pricing schemes and extra features preferences for mobility as a service: Sequential portfolio choice experiment. *Transportation Research Part A: Policy and Practice*, 131:123–148.

Cervero, R. and Duncan, M. (2008). Residential Self Selection and Rail Commuting: A Nested Logit Analysis.

Chakraborty, S., Rey, D., Levin, M. W., and Waller, S. T. (2021). Freeway network design with exclusive lanes for automated vehicles under endogenous mobility demand. *Transportation Research Part C: Emerging Technologies*, 133:103440.

Daly, A., Hess, S., Patruni, B., Potoglou, D., and Rohr, C. (2012). Using ordered attitudinal indicators in a latent variable choice model: a study of the impact of security on rail travel behaviour. *Transportation*, 39(2):267–297.

Daly, A. and Zachary, S. (1979). Improved multiple choice models. Teakfield: London.

Daziano, R. A. and Bolduc, D. (2013). Incorporating pro-environmental preferences towards green automobile technologies through a bayesian hybrid choice model. *Transportmetrica A: Transport Science*, 9(1):74–106.

Domencich, T. A. and McFadden, D. (1975). Urban travel demand. North-Holland Amsterdam.

Forinash, C. V. and Koppelman, F. S. (1993). Application and interpretation of nested logit models of intercity mode choice. *Transportation research record*, (1413).

Green, P. E. and Srinivasan, V. (1978). Conjoint analysis in consumer research: Issues and outlook. *Journal of Consumer Research*, 5(2):103–123.

Han, Y., Zegras, C., Pereira, F. C., and Ben-Akiva, M. (2020). A neural-embedded choice model: Tastenet-mnl modeling taste heterogeneity with flexibility and interpretability. *arXiv preprint arXiv:2002.00922*.

Hensher, D., Louviere, J., and Swait, J. (1998). Combining sources of preference data. *Journal of Econometrics*, 89(1-2):197–221.

Hess, S., Train, K. E., and Polak, J. W. (2006). On the use of a modified latin hypercube sampling (mlhs) method in the estimation of a mixed logit model for vehicle choice. *Transportation Research Part B: Methodological*, 40(2):147–163.

Hillel, T., Bierlaire, M., Elshafie, M., and Jin, Y. (2019). Weak teachers: Assisted specification of discrete choice models using ensemble learning. In *8th Symposium of the European association for research in transportation, Budapest*.

Hillel, T., Bierlaire, M., Elshafie, M. Z., and Jin, Y. (2021). A systematic review of machine learning classification methodologies for modelling passenger mode choice. *Journal of Choice Modelling*, 38:100221.

Hruschka, H., Fettes, W., Probst, M., and Mies, C. (2002). A flexible brand choice model based on neural net methodology a comparison to the linear utility multinomial logit model and its latent class extension. *OR spectrum*, 24(2):127–143.

Hunt, J. D. and Teply, S. (1993). A nested logit model of parking location choice. *Transportation Research Part B: Methodological*, 27(4):253–265.

Koppelman, F. S. (1976). Guidelines for aggregate travel prediction using disaggregate choice models. *Transportation Research Record*, 610:19–24.

Krueger, R., Bierlaire, M., Daziano, R. A., Rashidi, T. H., and Bansal, P. (2021). Evaluating the predictive abilities of mixed logit models with unobserved inter- and intra-individual heterogeneity. *Journal of Choice Modelling*, 41:100323.

Krueger, R., Rashidi, T. H., and Rose, J. M. (2016). Preferences for shared autonomous vehicles. *Transportation research part C: emerging technologies*, 69:343–355.

Krueger, R., Rashidi, T. H., and Vij, A. (2020). A dirichlet process mixture model of discrete choice: Comparisons and a case study on preferences for shared automated vehicles. *Journal of Choice Modelling*, page 100229.

Li, W. and Kamargianni, M. (2020). An integrated choice and latent variable model to explore the influence of attitudinal and perceptual factors on shared mobility choices and their value of time estimation. *Transportation science*, 54(1):62–83.

Liu, Y., Bansal, P., Daziano, R., and Samaranayake, S. (2019). A framework to integrate mode choice in the design of mobility-on-demand systems. *Transportation Research Part C: Emerging Technologies*, 105:648–665.

Louviere, J. J., Meyer, R. J., Bunch, D. S., Carson, R., Dellaert, B., Hanemann, W. M., Hensher, D., and Irwin, J. (1999). Combining sources of preference data for modeling complex decision processes. *Marketing Letters*, 10(3):205–217.

Luce, R. D. (1959). Individual choice behavior: A theoretical analysis, new york, ny: John willey and sons.

Marschak, J. (1960). Binary-choice constraints and random utility indicators. In *Mathematical Methods in the Social Sciences, 1959: Proceedings of the 1st Stanford Symposium*, number 4, pages 312–29. Stanford University Press.

McFadden, D. (1968). The revealed preferences of a public bureaucracy. *Dept. of Economics, Univ. of California, Berkeley*.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior.

McFadden, D. (1978). Modeling the choice of residential location. *Transportation Research Record*, (673).

McFadden, D. (1986). The choice theory approach to market research. *Marketing Science*, 5(4):275–297.

McFadden, D. (2001a). Disaggregate behavioral travel demand's rum side: A 30-year retrospective. pages 17–63.

McFadden, D. (2001b). Economic choices. *American Economic Review*, 91(3):351–378.

McFadden, D. and Reid, F. (1975). *Aggregate travel demand forecasting from disaggregated behavioral models*. Institute of Transportation and Traffic Engineering, University of California.

McFadden, D., Talvitie, A., Cosslett, S., Hasan, I., and Reid, F. (1977). Urban travel demand forecasting project phase i final report series, volume v. demand model estimation and validation.

Morikawa, T., Ben-Akiva, M., and Yamada, K. (1991). Forecasting intercity rail ridership using revealed preference and stated preference data. *Transportation Research Record*, 1328:30–35.

Motoaki, Y. and Daziano, R. A. (2015). A hybrid-choice latent-class model for the analysis of the effects of weather on cycling demand. *Transportation Research Part A: Policy and Practice*, 75:217–230.

Nuñez Velasco, J. P., Farah, H., van Arem, B., and Hagenzieker, M. P. (2019). Studying pedestrians' crossing behavior when interacting with automated vehicles using virtual reality. *Transportation Research Part F: Traffic Psychology and Behaviour*, 66:1–14.

Ortelli, N., Hillel, T., Pereira, F. C., de Lapparent, M., and Bierlaire, M. (2021). Assisted specification of discrete choice models. *Journal of Choice Modelling*, 39:100285.

Pacheco Paneque, M., Bierlaire, M., Gendron, B., and Sharif Azadeh, S. (2021). Integrating advanced discrete choice models in mixed integer linear optimization. *Transportation Research Part B: Methodological*, 146:26–49.

Paschalidis, E., Choudhury, C. F., and Hess, S. (2019). Combining driving simulator and physiological sensor data in a latent variable model to incorporate the effect of stress in car-following behaviour. *Analytic Methods in Accident Research*, 22:100089.

Paulssen, M., Temme, D., Vij, A., and Walker, J. L. (2014). Values, attitudes and travel behavior: a hierarchical latent variable mixed logit model of travel mode choice. *Transportation*, 41(4):873–888.

Paz, A., Arteaga, C., and Cobos, C. (2019). Specification of mixed logit models assisted by an optimization framework. *Journal of Choice Modelling*, 30:50–60.

Pereira, F. (2021). Rethinking travel behavior modeling representations through embeddings. *arXiv*, (arXiv:1909.00154).

Rodrigues, F., Ortelli, N., Bierlaire, M., and Pereira, F. C. (2020). Bayesian automatic relevance determination for utility function specification in discrete choice models. *IEEE Transactions on Intelligent Transportation Systems*.

Sifringer, B., Lurkin, V., and Alahi, A. (2020). Enhancing discrete choice models with representation learning. *Transportation Research Part B: Methodological*, 140:236–261.

Sivakumar, A., Bhat, C. R., and Ökten, G. (2005). Simulation estimation of mixed discrete choice models with the use of randomized quasi–monte carlo sequences: A comparative study. *Transportation Research Record*, 1921(1):112–122.

Small, K. and Rosen, H. S. (1981). Applied welfare economics with discrete choice models. *Econometrica*, 49:105–130.

Small, K. A., Winston, C., and Yan, J. (2005). Uncovering the distribution of motorists' preferences for travel time and reliability. *Econometrica*, 73(4):1367–1382.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological review*, 34(4):273.

van Cranenburgh, S., Wang, S., Vij, A., Pereira, F., and Walker, J. (2021). Choice modelling in the age of machine learning - discussion paper. *Journal of Choice Modelling*, page 100340.

Varotto, S. F., Farah, H., Toledo, T., van Arem, B., and Hoogendoorn, S. P. (2017a). Resuming manual control or not?: Modeling choices of control transitions in full-range adaptive cruise control. *Transportation Research Record*, 2622(1):38–47.

Varotto, S. F., Farah, H., Toledo, T., van Arem, B., and Hoogendoorn, S. P. (2018). Modelling decisions of control transitions and target speed regulations in full-range adaptive cruise control based on risk allostasis theory. *Transportation Research Part B: Methodological*, 117:318–341.

Varotto, S. F., Glerum, A., Stathopoulos, A., Bierlaire, M., and Longo, G. (2017b). Mitigating the impact of errors in travel time reporting on mode choice modelling. *Journal of Transport Geography*, 62:236–246.

Varotto, S. F., Jansen, R., Bijleveld, F., and van Nes, N. (2021). Driver speed compliance following automatic incident detection: Insights from a naturalistic driving study. *Accident Analysis & Prevention*, 150:105939.

Vij, A. and Krueger, R. (2017). Random taste heterogeneity in discrete choice models: Flexible nonparametric finite mixture distributions. *Transportation Research Part B: Methodological*, 106:76–101.

Vovsha, P. (1997). Application of cross-nested logit model to mode choice in tel aviv, israel, metropolitan area. *Transportation Research Record*, 1607(1):6–15.

Walker, J. L. (2001). *Extended discrete choice models: integrated framework, flexible error structures, and latent variables*. PhD thesis, Massachusetts Institute of Technology.

Walker, J. L., Ben-Akiva, M., and Bolduc, D. (2007). Identification of parameters in normal error component logit-mixture (neclm) models. *Journal of Applied Econometrics*, 22(6):1095–1125.

Walker, J. L. and Li, J. (2007). Latent lifestyle preferences and household location decisions. *Journal of Geographical Systems*, 9(1):77–101.

Wang, S., Mo, B., Hess, S., and Zhao, J. (2021). Comparing hundreds of machine learning classifiers and discrete choice models in predicting travel behavior: an empirical benchmark. *arXiv preprint arXiv:2102.01130*.

Williams, H. C. W. L. (1977). On the formation of travel demand models and economic evaluation measures of user benefit. *Environment and Planning A*, 9:285 – 344.

Wong, M. and Farooq, B. (2020). A bi-partite generative model framework for analyzing and simulating large scale multiple discrete-continuous travel behaviour data. *Transportation Research Part C: Emerging Technologies*, 110:247–268.

Wong, M. and Farooq, B. (2021). Reslogit: A residual neural network logit model for data-driven choice modelling. *Transportation Research Part C: Emerging Technologies*, 126:103050.

Zhao, X., Yan, X., Yu, A., and 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.