Journal of Choice Modelling 47 (2023) 100409

Contents lists available at ScienceDirect

Journal of Choice Modelling

journal homepage: www.elsevier.com/locate/jocm

Extensive hypothesis testing for estimation of mixed-Logit models

Prithvi Bhat Beeramoole [a], Cristian Arteaga [b], Alban Pinz [c], Md Mazharul Haque [a], Alexander Paz [a,*]

[a] School of Civil & Environment Engineering, Queensland University of Technology, Australia
[b] Howard R. Hughes College of Engineering, Civil and Environmental Engineering, University of Nevada-Las Vegas, Las Vegas, NV, USA
[c] Economic Research and Analysis, Queensland Department of Transport and Main Roads, Australia

ARTICLE INFO

Keywords:
Discrete outcome
Discrete choice
Bi-level optimization
Metaheuristic
Consumer behavior

ABSTRACT

Estimation of discrete outcome specifications involves significant hypothesis testing, including multiple modelling decisions which could affect results and interpretation. Model development is generally time-bound, and decisions largely rely on experience, knowledge of the problem context and statistics. There is often a risk of adopting restricted specifications, which could preclude important insights and valuable behavioral patterns. This study proposes a framework to assist in testing hypotheses and discovering mixed-Logit specifications that best capture discrete outcome behavior. The proposed framework includes a mathematical programming formulation and a bi-level constrained optimization algorithm to simultaneously test various modelling assumptions and produce meaningful specifications within a reasonable time. The bi-level framework illustrates the integration of a population-based metaheuristic with model estimation procedures. In addition, the optimization algorithm allows the analyst to impose assumptions on the models to test specific hypotheses or to ensure compliance with literature. Numerical experiments are conducted using different datasets and behavioral processes to illustrate the efficacy of the proposed extensive hypothesis testing in terms of interpretability and goodness-of-fit. Results illustrate the ability of the proposed algorithm to reveal important insights that can potentially be overlooked due to limited and/or biased hypothesis testing. In addition, the proposed extensive hypothesis testing generates multiple acceptable solutions, thereby suggesting potential directions for further investigation. The proposed framework can serve as a decision-assistance modelling tool in various applications, involving many variables and outcomes, such as road safety analysis, consumer choice behavior, and integrated land-use and travel choice models.

1. Introduction

1.1. Modelling discrete outcome problems

Discrete outcome models have been an integral part of behavioral studies, including but not limited to, transport and land-use planning, pathology analysis and market research. Even with the advent of advanced Machine Learning (ML) methods, discrete outcome models are still widely used because of their ability to capture behavior and estimate causality. Over the years, extensions and capabilities have been added to better capture behavior and explain influencing factors (Bierlaire, 1998). This development has led to a

* Corresponding author.
E-mail address: alexander.paz@qut.edu.au (A. Paz).

https://doi.org/10.1016/j.jocm.2023.100409
Received 29 March 2022; Received in revised form 15 February 2023; Accepted 21 February 2023
Available online 27 February 2023
1755-5345/

variety of modelling approaches, each with its own strengths and limitations (Mannering et al., 2016).

The process of developing discrete outcome specifications is highly involved as it requires an analyst to take several critical modelling decisions, including (1) selection of variables to be tested during model specification; (2) identification of variable forms and transformations (e.g., linear or non-linear); (3) variables to be tested with fixed or random coefficients; (4) distributional assumptions for random coefficients and error terms; and (5) selection of methods to deal with potential correlation, among others. Analysts generate candidate specifications or hypotheses and iteratively modify them until an acceptable model is obtained. The process is generally time-bound, often resulting in limited and slow testing of hypotheses, which can potentially introduce errors (Paz et al., 2019), leading to misspecifications, estimation biases, and erroneous predictions (Han et al., 2020).

The decisions during the specification process represent fundamental hypotheses, which can significantly affect results and interpretation of underlying behavior captured by models. Variables included in the model and their functional form represent hypotheses on how they influence a discrete outcome. Specifications assumed with a linear-in-parameters structure are generally a limited representation of reality (Kim et al., 2016). For example, empirical evidence suggests that travelers perceive a 10-min waiting time differently when travel time changes from 20 min to a few hours (Koppelman, 1981). A linear specification, however, imposes symmetrical behavior and is unable to capture potential empirical inflection point(s) showing sudden shift(s) in preferences (Mandel et al., 1997). Popular methods to capture nonlinearity in discrete outcome models include variable transformations, such as logarithmic (Ben-Akiva et al., 1987), hyperbolic (Kitazawa, 2012), exponential (Gaudry, 1981), and piecewise regression (Kim et al., 2014). Although these methods have been found to capture more information on causality compared to linear specifications, the type of transformation and the associated variables need to be predetermined. Some studies have applied Box-Cox transformations (Box and Cox, 1964) to approximate a linear, logarithmic, or exponential effect depending on an estimated transformation coefficient. The Box-Cox relieves the analyst from predefining a transformation, thereby alleviating misspecification to some extent (Kim et al., 2016; Orro et al., 2005). However, the selection of variables to be tested with a transformation still largely relies on the analyst.

Studies have demonstrated a significant improvement in terms of goodness-of-fit and explanatory power when random coefficients are included in the specification (Anowar and Eluru, 2018; Sillano and Ortúzar, 2005; Train, 2003). While statistical tests are available (Fosgerau and Hess, 2009), the analyst needs to perform a series of analyses to determine variables to be estimated with random coefficients along with the mixing distribution. Hence, overall model results, including interpretability and goodness-of-fit, should be used to select distributions. Hensher and Greene (2003) tested the effect of different distributional assumptions in a transportation mode choice study. They found that a proportion of the sample with a negative value of travel-time savings changed depending on the assumed distribution from 19.21% to 37.92% and 39.33% for normal to uniform and triangular distributions, respectively. These results suggest that a data-driven approach involving testing multiple hypothesis is essential to support the corresponding modelling decisions and get the best possible interpretation and goodness-of-fit. Alternative approaches using flexible distributions have also been recommended to alleviate the need for testing multiple random distributions for an adequate representation of unobserved heterogeneity. Examples include those proposed by Train (2008); Train (2016); and Keane and Wasi (2013), and non-parametric and semi-parametric distributions that are used in latent class models (Greene and Hensher, 2003; Vij and Krueger, 2017). However, a careful consideration of the trade-offs between model flexibility and data overfitting is required to support the specification selection (Fosgerau and Bierlaire, 2007).

Similarly, multiple sources of correlation can exist within the data depending upon the study context (Hess and Rose, 2012). Crucial information could be lost when a restricted covariance structure is imposed a priori on the specification. Studies had also reported a change in coefficient signs when variables were allowed to correlate, indicating potential misspecification issues if the correlation is unaccounted for (Fountas et al., 2018). Based on a detailed review, Hess and Train (2017) recommend testing different restrictions on the covariance matrix and selecting an appropriate approach.

Although there is vast literature to provide guidance to test flexible model specifications for non-linearities, random and correlated random parameters, there is limited opportunity and resources for analysts to test large numbers of associated hypotheses and methods. Hence, restricted specifications with limited capabilities to address modelling challenges are frequently adopted (Fiebig et al., 2010; Paz et al., 2019; Train, 2007; Walker and Ben-Akiva, 2002). Further, the availability of highly dimensional datasets often makes the analysis more laborious and challenging. Considering that problem size grows substantially as data dimensionality increases (Vinterbo and Ohno-Machado, 1999), an exhaustive hypothesis testing searching for a solution that addresses all data and modelling aspects is not feasible (Fan and Li, 2006).

1.2. Relevant studies and proposed contribution

Seeking to develop alternatives to statistical-based models and address some of the above modelling challenges, several applications of Machine Learning (ML) can be found in the discrete-outcome modelling literature, including Neural Networks (Dia and Panwai, 2010; Ma et al., 2020; Ramsey and Bergtold, 2020; Sifringer et al., 2018; Wang et al., 2020a; Wang et al., 2020b), Decision Trees (Brathwaite et al., 2017), Random Forests (Wang and Kim, 2019; Zhao et al., 2020), and Support Vector Machines (Martín-Baos et al., 2021). The ML-based approaches typically outperform predictions of standard statistical methods as they utilize complex model specifications. However, standard measures to interpret behavior, such as estimation of willingness to pay (WTP), have been found inconsistent with ML-based models (Ramsey and Bergtold, 2020; Zhao et al., 2020). Alternative approaches for estimating substitution patterns, including numerical approximation (Zhao et al., 2020), such as gaussian quadrature methods (Ramsey and Bergtold, 2020) have been proposed. However, case-specific methods need to be applied based on the ML technique, while computational cost required by such methods is often high. In addition, there are concerns regarding potential over-fitting and estimation issues associated with hyperparameters required by ML-based approaches (Han et al., 2020; Martín-Baos et al., 2021).

Integrated approaches have also been proposed to maintain behavioral interpretability, while aiming to alleviate modelling decisions. Brathwaite et al. (2017) developed a Bayesian decision tree framework to model semi-compensatory choice behavior. The study developed a method using Bayesian trees to link ML-based techniques with random utility theory. However, the framework is unable to consider a combination of behavioral features, such as nonlinearity and heterogeneity (Brathwaite et al., 2017). Hybrid specifications have also been proposed in which utility is partly prespecified by the analyst and partly estimated using ML techniques (Sifringer et al., 2018; Sifringer et al., 2020). Han et al. (2020) added a neural network component to a manual specification embedded to model heterogeneity in choice preferences. Semi-nonparametric approaches have also been applied to preserve the validity of standard WTP measures. The objective is to ease the distributional assumptions of the error term using compound functions, such as flexible Fourier forms (Creel and Loomis, 1997) and Hermite-polynomial methods (Arouna and Dabbert, 2012). Integrated approaches coupling statistical modelling with ML techniques have shown improved accuracy in predictions along with some behavioral interpretation (van Cranenburgh et al., 2021). However, the proposed approaches do not assist the analyst in making critical decisions required for model development. In addition, model performance was compared with standard restrictive specifications. Therefore, potential trade-offs of using the integrated approaches over specifications that simultaneously consider nonlinear, heterogeneous, and correlated effects are yet to be investigated in detail.

Studies have used data-driven metaheuristics to assist discrete outcome model development while testing hypotheses (Ortelli et al., 2021; Rodrigues et al., 2019). The general idea is to engage an impartial mechanism to explore multiple model structures and best-fit combinations of features from the dataset to include them in the specification. Paz et al. (2019) used simulated-annealing (SA) for developing mixed-Logit models with random parameters while selecting potential explanatory variables and the corresponding densities for their coefficients. Ortelli et al. (2021) adopted a variable neighborhood search method to generate discrete choice specifications. Rodrigues et al. (2019) used automatic relevance determination in a Bayesian framework to identify important features from the data by restricting a preference parameter for irrelevant features to zero. These data-driven approaches were found to significantly reduce analysts' efforts while preserving causal interpretability. However, each of the previous studies focused on specific features of the model development process. Both Ortelli et al. (2021) and Rodrigues et al. (2019) contributed to the identification of important variables and their functional forms but limited the search to multinomial and nested logit models. In contrast, Paz et al. (2019) generated mixed-Logit models using a linear-in-parameters specification and uncorrelated coefficients. While the methods proposed by Paz et al. (2019) and Ortelli et al. (2021) identified adequate specifications with improved goodness-of-fit relative to benchmark solutions from the literature, a detailed analysis to illustrate the significance of an optimization framework to perform an extensive hypothesis testing to capture important behavioral patterns is not available.

This study seeks to move towards the development of an unbiased and efficient framework to perform extensive hypothesis testing considering simultaneously multiple modelling decisions, including potential explanatory variables, the type of parameters to be estimated (outcome-specific or generic), presence of unobserved heterogeneity, non-linearity, and correlations. Simultaneous consideration of these modelling decisions significantly increases problem complexity and the required hypothesis testing. Previous studies have attempted to develop frameworks to assist with critical modelling decisions, while preserving behavioral realism (Ortelli et al., 2021; Paz et al., 2019). This study adds to the literature by developing an efficient optimization framework to simultaneously test many modelling assumptions and assist in discovering behavioral insights beyond those often reported in the literature using restricted specifications. A bi-level optimization approach is proposed to perform the required extensive hypothesis testing. The proposed framework enables the investigation and estimation of specifications for a large range of applications, including those involving many variables, outcomes, and behavioral nuances.

The remaining of this manuscript is organized as follows. Section 2 presents the proposed methodological framework, including the problem formulation and extensive hypothesis testing framework. Section 3 provides numerical experiments conducted using the proposed framework and comparative analyses with existing methods. Finally, section 5 delivers conclusions and identifies future research directions. Throughout this paper, the words "specification" and "solution" are used interchangeably.

2. Methodology

A mathematical programming formulation is presented below including an objective function, decision variables and associated constraints. Similarly, a metaheuristic-based solution algorithm is developed, implemented, and tested to search for adequate mixed-Logit specifications. The proposed framework provides flexibility to include any a priori knowledge about data or problem context that can guide the hypothesis testing. For example, level-of-service attributes such as travel time and cost are typically important aspects while estimating travel demand. Analysts would often require testing of such specific effects irrespective of their statistical significance. Flexibilities are, therefore, essential to align the specification search with the study objectives.

2.1. Mathematical programming problem

2.1.1. Notation and definitions

This subsection introduces notation and definitions used to formulate the proposed mathematical programming problem and solution algorithm for extensive hypothesis testing during the estimation of mixed-Logit models.

Data Inputs	
Set	
$\overline{N}$	Set of individual observations; $\overline{N} = \{1,\ldots,N\}$; indexed on $n \in \overline{N}$
$\overline{J}$	Set of discrete outcomes; $\overline{J} = \{1,\ldots,J\}$; indexed on $j \in \overline{J}$
$\overline{T}$	Set of time events when observations n were collected; $\overline{T} = \{1,\ldots,T\}$; indexed on $t \in \overline{T}$
$\overline{K}$	Set of potential alternative-specific attributes; $\overline{K} = \{1,\ldots,K\}$; indexed on $k \in \overline{K}$
$\overline{M}$	Set of potential characteristics associated with individual observations; $\overline{M} = \{1,\ldots,M\}$; indexed on $m \in \overline{M}$
F	Set of possible probability density functions considered for the estimation of coefficients; $F = \{\text{lognormal}, \text{normal}, \text{uniform}, \text{truncated normal}, \text{triangular}\}$
Index	
l	Subscript to denote a base outcome; $l \in \overline{J}$
p	Subscript to denote an alternative-specific attribute; $p \in \overline{K}$
Observed Variables	
$\mathbf{x}_{nj}^{t}$	Vector containing measurements of alternative-specific attributes $\overline{K}$ for alternatives considered by individual n in $t \forall k \in \overline{K}$ and $j \in \overline{J}$; $\mathbf{x}_{nj}^{t} = [x_{nj1}^{t},\ldots,x_{njk}^{t},\ldots,x_{njK}^{t}]$
$\mathbf{z}_n$	Vector containing measurements of observation-specific characteristic m for individual observation $n \forall m \in \overline{M}$; $\mathbf{z}_n = [z_{n1},\ldots,z_{nm},\ldots,z_{nM}]$
y_{nj}^{t}	Binary variable taking value 1 if j is the observed outcome n at t; 0 otherwise
Pre-specifications	
$\widehat{\alpha}_{jk}$	Binary variable taking value 1 if x_{njk}^{t} is pre-specified to be included in the model; 0 otherwise
$\widehat{\ddot{\alpha}}_{jm}$	Binary variable taking value 1 if z_{nm} is pre-specified to be included in the specification for alternative j; 0 otherwise
$\daleth_{k,m}$	Binary variable taking value 1 if a variable interaction is pre-specified between x_{njk}^{t} and z_{nm}; 0 otherwise
$\widehat{g}(z_{nm}, x_{njk}^{t})$	Pre-specified functional form for interaction between x_{njk}^{t} and z_{nm}
$\widehat{\omega}_k$	Binary variable taking value 1 if transformation is pre-specified for x_{njk}^{t}; 0 otherwise
μ_k	Binary variable taking value 1 if λ_k is pre-specified for x_{njk}^{t}; 0 otherwise
$\widehat{\lambda}_k$	Pre-specified value of λ_k for x_{njk}^{t}
γ_k	Binary variable taking value 1 if f_k is pre-specified; 0 otherwise
$\widehat{f}_k$	Pre-specified random distribution; $\widehat{f}_k \in F$
$\widehat{\varphi}_{k,p}$	Binary variable taking value 1 if correlation between β_{njk} and β_{njp} is pre-specified; 0 otherwise $\forall n \in \overline{N}$; $k, p \in \overline{K}$ and $p \neq k$
Decision variables	
$\boldsymbol{\alpha}_j$	Vector containing binary variables α_{jk} taking value 1 if x_{njk}^{t} is included in the specification for j; 0 otherwise $\forall j \in \overline{J}$ *and* $k \in \overline{K}$; $\boldsymbol{\alpha}_j = [\alpha_{j1},\ldots\alpha_{jk},\ldots,\alpha_{jK}]$
$\ddot{\boldsymbol{\alpha}}_j$	Vector containing binary variables $\ddot{\alpha}_{jm}$ taking value 1 if z_{nm} is included in the specification for alternative j $\forall m \in \overline{M}$; 0 otherwise; $\ddot{\boldsymbol{\alpha}}_j = [\ddot{\alpha}_{11},\ldots\ddot{\alpha}_{jm},\ldots,\ddot{\alpha}_{jM}]$
ω_k	Binary variable taking value 1 if a variable transformation is applied on x_{njk}^{t}; 0 otherwise
$\boldsymbol{\lambda}$	Vector of coefficients used to determine a transformation for variables in $\mathbf{x}_{nj}^{t}$; ; $\boldsymbol{\lambda} = [\lambda_1,\ldots,\lambda_k,\ldots,\lambda_K]$
$\boldsymbol{\beta}_n$	Matrix of size $N \times J \times K$ containing coefficients for $\mathbf{x}_{nj}^{t} \forall k \in \overline{K}$ and $j \in \overline{J}$; $\boldsymbol{\beta}_n = [\beta_{n11},\ldots,\beta_{njk},\ldots,\beta_{nJK}]$
$\mathbf{f}$	Vector of probability density functions for $\boldsymbol{\beta}_n$; $\mathbf{f} = [f_1,\ldots,f_k,\ldots,f_K], \forall f_k \in \mathrm{F}$
$\boldsymbol{\theta}_j$	Vector of coefficients for $\mathbf{z}_n$ for alternative j ; $\boldsymbol{\theta}_j = [\theta_{j1},\ldots,\theta_{jm},\ldots,\theta_{jM}]$
β_{jk}	Location parameter of the random distribution corresponding to β_{njk}
σ_{jk}	Scale parameter of the random distribution corresponding to β_{njk}
$\sigma_{jk,jp}$	Cholesky coefficients for the covariance matrix $\boldsymbol{\Sigma}$ of $\boldsymbol{\beta}$ $\forall k, p \in \overline{K}$
$\boldsymbol{\Gamma}$	Matrix of size $J \times K$ with elements $\tau_{jk}, 1 \leq \tau_{jk} \leq J$, which are indicators enabling the estimation of β_{njk} as a generic or outcome-specific coefficients
$\varphi_{k,p}$	Indicator variable taking value 1 $\Longleftrightarrow \beta_{njk}$ *and* β_{njp} are correlated; 0 otherwise $\forall k, p \in \overline{K}$ and $p \neq k$
δ	Number of estimable parameters in the specification

2.1.2. Problem formulation

The observed utility associated with outcome j and individual n in observation t is given by v_{nj}^{t} eqn. (1). Depending on the contribution to fit and intuitive behavioral meaning, β_{njk} can be estimated as a fixed, random, or random-correlated coefficients, as given by eqn. (2). For observation-specific characteristics $\mathbf{z}_n$, such as socio-demographic variables or household attributes used in discrete choice analysis, the corresponding coefficients θ_{jm} are estimated as alternative-specific. Eqn. (3) is used to normalize the base-alternative coefficients to 0, ensuring that a maximum of $J-1$ coefficients are estimated for the given observation-specific characteristic.

$$v_{nj}^{t} = \ddot{\boldsymbol{\alpha}}_j \boldsymbol{\theta}_j \mathbf{z}_n + \boldsymbol{\alpha}_j \boldsymbol{\beta}_n {\mathbf{x}_{nj}^{t}}^{(\boldsymbol{\lambda})} \tag{1}$$

$$\beta_{njk} = \begin{cases} \beta_{jk} \quad \textit{if } \beta_{njk} \text{ estimated as fixed coefficient} \forall n \text{ ;} \\ f_k(\beta_{jk}, \sigma_{jk}) \quad \textit{if } \beta_{njk} \text{ estimated as random uncorrelated coefficient;} \\ f_k(\beta_{jk}, \sigma_{jk,jp}) \quad \textit{if } \beta_{njk} \text{ estimated as random correlated coefficient.} \end{cases} \tag{2}$$

$$\theta_{jm} = \begin{cases} 0 & \textit{if } j = l \forall m \in \overline{M}; \\ \theta_{jm} & \textit{otherwise} \end{cases} \tag{3}$$

The extensive hypothesis testing required to estimate mixed-Logit models is viewed in this study as a bi-level non-linear mixed-integer optimization problem. The **lower-level objective function** is to maximize the log-likelihood estimate (LL) given by eqn. (4) , including the following **decision variables ($\boldsymbol{D}$)**:

(i) coefficients $\beta_{njk} \forall \alpha_{jk} = 1$,
(ii) coefficients $\theta_{jm} \forall \ddot{\alpha}_{jm} = 1$,
(iii) coefficients λ_k to determine a transformation for $x^t_{njk} \forall \alpha_{jk} = \omega_k = 1$,
(iv) coefficients $\sigma_{k,p}$ $\forall \alpha_{jk} = \alpha_{jp} = \varphi_{k,p} = 1$.

The log-likelihood estimate is defined using the generalized linear modelling framework with an unobserved component of utility assumed to be i.i.d extreme value. In the present form, the proposed formulation can be applied to estimate multinomial and mixed-Logit models, while investigating for nonlinear interactions, heterogeneity, and correlated effects. Mixed-Logit models are highly flexible and can be approximated to any random utility model (McFadden and Train, 2000), when specified correctly including adequate distributions. Mixed-Logit models overcome the three fundamental limitations of standard Logit models by allowing preference heterogeneity, unrestricted substitution patterns and correlation in unobserved factors over time (Train, 2003). In addition, the approximate maximum likelihood estimations using simulation methods have further facilitated mixed-Logit applications.

A straightforward extension of the formulation to estimate nested Logit models can also be achieved by including constraints to define nesting structures. Furthermore, the lower-level objective function can be reformulated using other modelling approaches, including Bayesian methods, which use deviance under the posterior estimates (Geedipally et al., 2014), or ordinary least squares (OLS) which uses the sum of squared errors (Khadka et al., 2018), and global maximum likelihood estimation methods (Liu and Mahmassani, 2000).

$$Max.LL = \sum_{n=1}^{N} \int \left\{ \prod_{t=1}^{T} \prod_{j=1}^{J} \left(\frac{e^{\ddot{\boldsymbol{\alpha}}_j \boldsymbol{\theta}_j \mathbf{z}_n + \boldsymbol{\alpha}_j \boldsymbol{\beta}_n \mathbf{x}^{t\,(\boldsymbol{\lambda})}_{nj}}}{\sum_{j=1}^{J} e^{\ddot{\boldsymbol{\alpha}}_j \boldsymbol{\theta}_j \mathbf{z}_n + \boldsymbol{\alpha}_j \boldsymbol{\beta}_n \mathbf{x}^{t\,(\boldsymbol{\lambda})}_{nj}}} \right)^{y^t_{nj}} \right\} \mathbf{f}(\boldsymbol{\beta}_n | \boldsymbol{\beta}, \boldsymbol{\Sigma}) \mathbf{d}\boldsymbol{\beta}_n \tag{4}$$

A critical aspect during model development is having a selection criterion or objective function that enables to achieve the desired outcomes. For the estimation of a large range of models, both the Akaike Information Criteria (AIC) (Akaike, 1998) and Bayesian Information Criteria (BIC) (Schwarz, 1978) have provided better results than log-likelihood, because they include a penalizing factor based on the number of model parameters. The BIC has provided stability in problems involving a large number of observations and modelling parameters (Wu et al., 2020). Similarly, the BIC was selected as the objective function to ensure that quality models were generated during the search while maintaining parsimony (Khadka and Paz, 2017; Khadka et al., 2018; Veeramisti et al., 2021). Paz et al. (2019) used the BIC, whereas Ortelli et al. (2021) proposed a multi-objective framework, including maximization of the log-likelihood and minimization of the number of model parameters. Eqn. (5) provides the general form of AIC and BIC (Román et al., 2017; Vrieze, 2012):

$$AIC \text{ or } BIC = -2LL(\boldsymbol{\beta}) + \eta\rho \tag{5}$$

where $LL(\boldsymbol{\beta})$ is the log-likelihood estimate of a model with ρ parameters, η is a penalty coefficient taking value 2 for AIC and log of number of observations for BIC. The penalty coefficient offers a check on overfitting while selecting an adequate subset of predictor variables to explain an outcome. Although none of these criteria guarantees the identification of optimal solutions (Parady et al., 2021), BIC has been widely used in discrete outcome modelling due to its heavier penalty coefficient, which enables the identification of specifications with relatively better estimates using lesser model parameters (Ortelli et al., 2020; Román et al., 2017; Vij et al., 2018). In this study, the **upper-level objective function** is to minimize the BIC given by eqn. (6), including the following **decision variables**, which define a specification ($\boldsymbol{M}$):

(i) α_{jk} indicating whether variable x^t_{njk} is included in the specification $\forall\, j \in \overline{J}, k \in \overline{K}$,
(ii) $\ddot{\alpha}_{jm}$ in dicating whether variable z_{nm} is included in the specification $\forall\, m \in \overline{M}$,
(iii) ω_k indicating whether a transformation is applied to $x^t_{njk} \forall j \in \overline{J}, k \in \overline{K}$,
(iv) density function f_k for coefficient β_{jk} $\forall\, f \in F, j \in \overline{J}, k \in \overline{K}$,
(v) τ_{jk} indicating whether β_{jk} is estimated as a generic or outcome-specific coefficients $\forall j \in \overline{J}, k \in \overline{K}$, and
(vi) $\varphi_{k,p}$ indicating whether β_{jk} *and* β_{jp} are correlated $\forall\, f \in F, j \in \overline{J}, k, p \in \overline{K}$ and $p \neq k$.

$$\text{Min. BIC} = \delta \ln(N) - 2 * \left[\sum_{n=1}^{N} \int \left\{ \prod_{t=1}^{T} \prod_{j=1}^{J} \left(\frac{e^{\ddot{\boldsymbol{\alpha}}_j \boldsymbol{\theta}_j \mathbf{z}_n + \boldsymbol{\alpha}_j \boldsymbol{\beta}_n \mathbf{x}_{nj}^{t\ (\boldsymbol{\lambda})}}}{\sum_{j=1}^{J} e^{\ddot{\boldsymbol{\alpha}}_j \boldsymbol{\theta}_j \mathbf{z}_n + \boldsymbol{\alpha}_j \boldsymbol{\beta}_n \mathbf{x}_{nj}^{t\ (\boldsymbol{\lambda})}}} \right)^{y_{nj}^t} \right\} \mathbf{f}(\boldsymbol{\beta}_n | \boldsymbol{\beta}, \boldsymbol{\Sigma}) \mathbf{d}\boldsymbol{\beta}_n \right] \tag{6}$$

subject to:

$$\alpha_{jk}, \ddot{\alpha}_{jm}, \omega_k, \varphi_{k,p,} \widehat{\alpha}_{jk}, \widehat{\ddot{\alpha}}_{jm}, \widehat{\omega}_k, \mu_k, \gamma_k, \widehat{\varphi}_{k,p}, \beth_{m,k} \in \{0, 1\} \forall m, j, k, and\ p, p \neq k \tag{7}$$

$$x_{njk}^{t\ (\lambda_k)} = \begin{cases} \dfrac{x_{njk}^t - 1}{\lambda_k} & if\ x_{njk}^t \in \mathbb{R}, \omega_k = 1\ and\ \lambda_k \neq 0 \forall x_{njk}^t\ ; \\ \ln x_{njk}^t & if x_{njk}^t \in \mathbb{R}^+, \omega_k = 1\ and\ \ \lambda_k = 0 \forall x_{njk}^t; \\ \quad x_{njk}^t & if\ \omega_k = 0 \forall x_{njk}^t \end{cases} \tag{8}$$

$$If\ \tau_{jk} = \tau_{rk}, then\ \beta_{jn} = \beta_{rn} \forall j, r \in J\ and\ j \neq r \tag{9}$$

$$If\ \widehat{\alpha}_{jk} = 1, then\ \alpha_{jk} = 1 \forall j\ and\ k \tag{10}$$

$$If\ \widehat{\ddot{\alpha}}_{jm} = 1, then\ \ddot{\alpha}_{jm} = 1 \forall m \tag{11}$$

$$If\ \widehat{\omega}_k = 0, then \omega_k = 0 \forall k \tag{12}$$

$$if\ \widehat{\omega}_k = 1\ and \mu_k = 0, then \omega_k = 1 \forall k \tag{13}$$

$$if\ \widehat{\omega}_k = \mu_k = 1, then\ \omega_k = 1\ and\ \lambda_k = \widehat{\lambda}_k \forall k \tag{14}$$

$$If\ \beth_{m,k} = 1, then\ x_{njk'}^t = \widehat{g}\left(z_{nm}, x_{njk}^t\right) \forall j, m\ and\ k, and\ k' \neq k \tag{15}$$

$$if\ \gamma_k = 1, then\ f_k = \widehat{f}_k \forall k \tag{16}$$

$$if\ \widehat{\varphi}_{k,j} = 1, then\ \varphi_{k,p} = 1 \forall k\ and\ p, p \neq k \tag{17}$$

Constraints (7) ensure that variables $\alpha_{jk}, \ddot{\alpha}_m, \omega_k, \varphi_{k,p,} \widehat{\alpha}_{jk}, \widehat{\ddot{\alpha}}_m, \widehat{\omega}_k, \mu_k, \gamma_k, \widehat{\varphi}_{k,p}, \beth_{m,k}$ only take values zero or one to test a particular hypothesis in the specification as provided in their definition under section 2.1.1. Constraints (8) are used to capture a generalized effect of variables on outcomes as proposed by Bierlaire (1998); Orro et al. (2005). Coefficient λ_k applies a Box-Cox transformation on x_{njk}^t by testing linear, logarithmic, and power functional forms.

Constraints (9) impose the estimation of β_{jk} either as generic or outcome-specific for all x_{njk}^t. If all values in the k^{th} column of Γ are unique, alternative-specific coefficients are estimated for the corresponding k^{th} explanatory variable. Conversely, if the k^{th} column of Γ have the same values, a generic coefficient is estimated. For example, if $\tau_{1k} \neq \tau_{2k}$, then alternative-specific β_{1k} and β_{2k} will be estimated, but if $\tau_{1k} = \tau_{2k}$, a generic coefficient $\beta_{1k} = \beta_{2k}$ will be estimated for alternatives $j = [1, 2]$.

Pre-specifications are imposed using constraints (10)-(17) to include any a priori knowledge into the model development. Constraints (10) and (11) impose pre-specified variables into the specification. Constraints (12) ensure that x_{njk}^t, if included in the specification, enters as linear-in-parameters. Constraints (13) ensure that the transformation of x_{njk}^t is determined by estimating λ_k. Constraints (14) impose a pre-specified transformation on x_{njk}^t using $\widehat{\lambda}_k$. A new variable $x_{njk'}^t$ can be defined based on a pre-specified functional form between x_{njk}^t and z_{nm} using constraints (15) to represent an interaction effect between specific individual characteristics and alternative attributes that may be essential in some behavioral analysis (Espino et al., 2006). For example, travel demand analysis has used the interaction of travel cost and income, to analyze the sensitivity of behavior to travel budget (Blaine et al., 2015). In addition, the coefficient distribution (constraints (16)) and/or estimation of correlated coefficients (constraints (17)) can be included or tested in the specification even when the associated statistics do not necessarily warrant them. These pre-specifications ensure that the generated models align with the problem objectives and enable the consideration of important practical aspects beyond the statistics as often required in causal analyses.

The proposed mathematical programming problem illustrates how multiple modelling decisions can be considered simultaneously for an extensive hypothesis testing. However, the proposed problem formulation does not include testing all random parameter distributions, nor all forms of interactions and nonlinear transformations. Additional constraints and the corresponding search capabilities in the solution algorithm are required to consider other modelling decisions such as testing flexible distributions including mixture-of-normal (Fosgerau and Hess, 2009; Keane and Wasi, 2013), nonparametric finite mixture distributions (Vij and Krueger, 2017), and other forms of variable interactions and nonlinear transformations, for instance piecewise linear approximations (Ben-Akiva et al., 1985).

2.2. Extensive hypothesis testing algorithm

A metaheuristic algorithm is proposed to solve the above mathematical programming problem. The upper level is a mixed-integer programming problem involving highly dimensional and non-linear hypotheses testing (search space), including potential explanatory variables, parameter type (outcome-specific or generic), unobserved heterogeneity, transformations, and correlations. The upper-level problem is solved using an improved global-best harmony search (IGBHS) (Xiang et al., 2014) to generate and test potential model specifications. Harmony search is a population-based metaheuristic inspired by music improvisation processes (Geem et al., 2001), wherein musicians improve initial harmonies by iteratively adjusting their pitch.

Harmony search and its variants have been effectively used to solve relevant optimization problems involving many decision variables such as the calibration of traffic flow simulation models (Cobos et al., 2020), selection of optimal routes and frequencies for bus rapid transit systems (Ruano-Daza et al., 2018), identifying optimal signal settings for a road transport network (Ceylan and Ceylan, 2012), and optimizing production and scheduling for a supply chain management problem (Guo et al., 2017). The IGBHS has been adopted in this study because of its distinctive capabilities, including (1) multiple start points to escape potential local optima (Diao and Shen, 2012), (2) utilization of multiple search strategies such as opposition-based learning and greedy-based search approaches, (3) exploitation near potential solutions, (4) low sensitivity to changes and relatively simple tuning of hyperparameters (Alia and Mandava, 2011; Kattan et al., 2010), and (5) ability to conduct an effective search within pre-set constraints.

A single best specification may not exist for a given problem unless it is for a synthetically generated dataset. The IGBHS stores multiple candidate specifications after suboptimal solutions undergo improvisation. Therefore, alternative specifications are available for selection and further improvisation to obtain "most suitable and parsimonious" solution. It is important to emphasize that the objective is not to replace or substitute the analyst but rather facilitate extensive hypotheses testing within a reasonable time.

The lower-level solution, whose estimates provide inputs to evaluate the upper-level objective function, is illustrated in Fig. 1. A specification or solution M, containing vectors of decision variables generated in the upper level, is used as an input to solve the lower-level problem. A simulated maximum likelihood estimation (MLE) procedure involving a gradient-based search is used to solve the lower-level problem, including the corresponding decision variables D. Search and evaluation continues until all decision variables, which were not pre-specified, are statistically significant according to a desired significance level p^*.

The algorithm utilized to solve the entire specification problem is illustrated in Fig. 2. A critical aspect that affects the performance of most solution algorithms is the fine-tuning of hyperparameters (Emaasit and Paz, 2018). In this study, hyperparameters were defined in the 'Initialization' step using covering arrays. These values include harmony memory size (HMS), minimum and maximum harmony memory consideration rate ($HMCR_{min}$ *and* $HMCR_{max}$), minimum and maximum pitch adjustment rate (PAR_{min} and PAR_{max}), maximum number of iterations ($iter_{max}$), proportion of iterations to be completed before initiating local search (ρ), and threshold to compare new solutions with those in HM (Δ).

A harmony memory (HM) of size HMS is initiated with randomly generated specification M. Known or anticipated specifications, such as the ones in Stated Preference experiments, can also be included in HM to save time and/or test improvements and their quality. An opposition-based learning (OBL) algorithm (Rahnamayan et al., 2008) is then implemented to initialize an opposite harmony memory (OHM). For each specification M in HM, an opposite model specification (OM) is generated using features that were not included in M to ensure extensive hypothesis testing (search). The two sets of random specifications provide multiple start points for

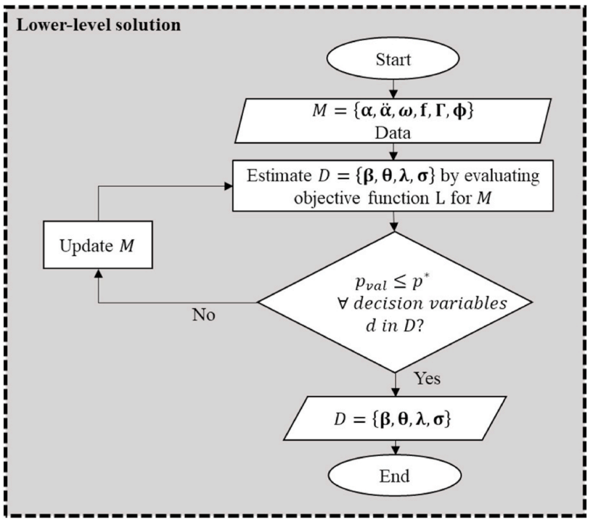

Fig. 1. Flowchart illustrating the algorithm used to solve the lower level of the proposed mathematical programming problem.

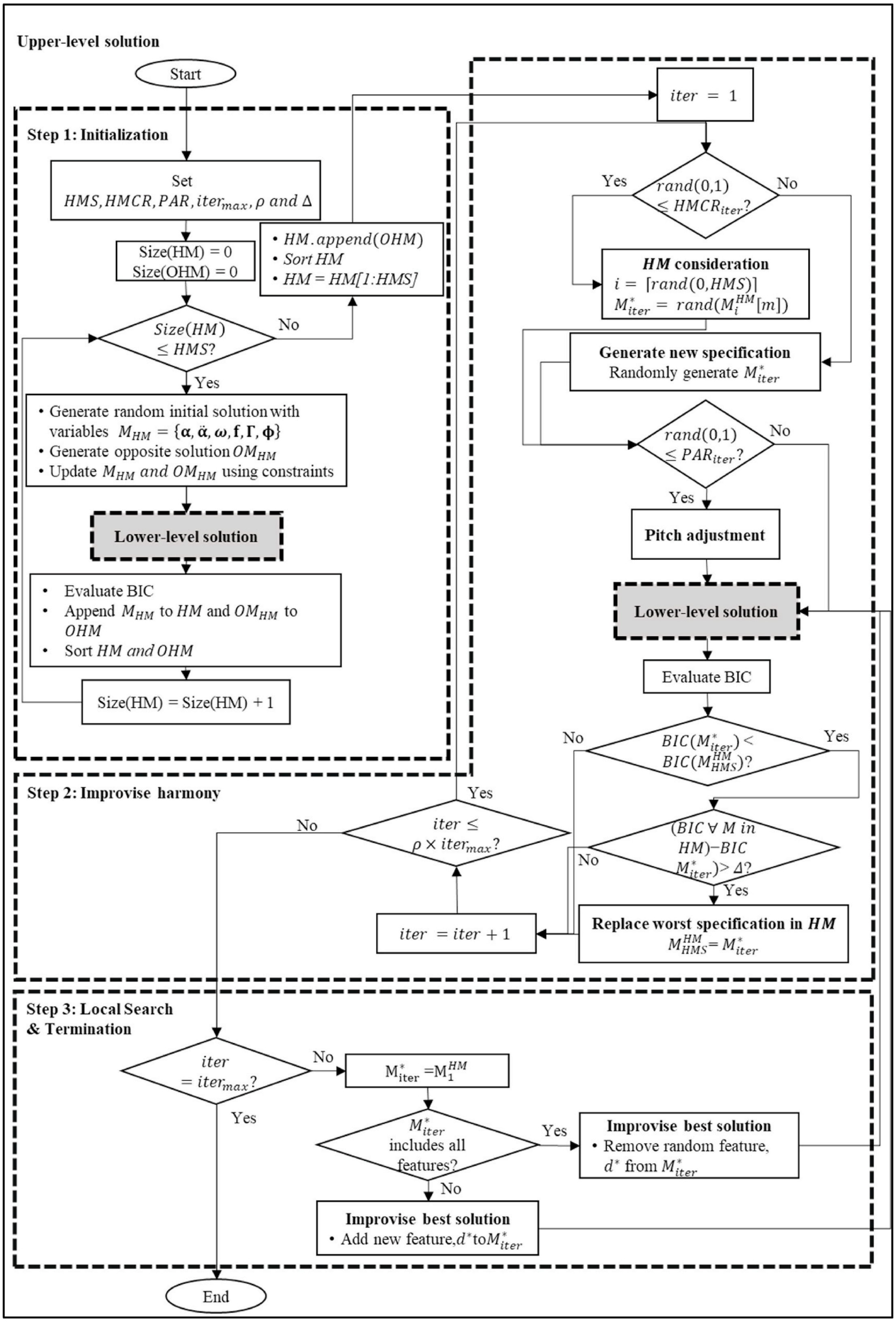

Fig. 2. Flowchart illustrating the algorithm used to solve the proposed mathematical programming problem.

the search and significantly improve exploration. The two memories are then combined to generate an initial harmony memory of size *HMS* after sorting the specifications based on their goodness-of-fit.

An 'Improvise harmony' step is initiated where the solutions in memory are perturbed based on dynamic values of *HMCR* and *PAR* given by eqns. (18) and (19), respectively (Xiang et al., 2014). The use of dynamic *HMCR* and *PAR* has been proposed to avoid premature convergence. A new solution is created either by considering previous specifications in memory or by generating a new combination of decision variables, representing a new hypothesis to test. The pitch adjustment step follows wherein the specification is fine-tuned by adding or removing some features. At each of these steps, the new specification is tested against the worst in memory, which is replaced if a better solution is found.

$$HMCR_{iter} = \left[HMCR_{min} + \left(\frac{HMCR_{max} - HMCR_{min}}{iter_{max}}\right) iter\right] \tag{18}$$

$$PAR_{iter} = \left[PAR_{min} + \left(\frac{PAR_{max} - PAR_{min}}{iter_{max}}\right) iter\right] \tag{19}$$

The local search step is initiated towards the final stages when iterations reach a pre-defined threshold. A greedy-based strategy is deployed, wherein the best specification in memory is exploited to seek a better fit. For every change in the feature combination, the objective function is evaluated to check for an improvement in fit. If the new solution is unique and better than any other solution in memory by a tolerance value Δ, the worst solution in *HM* is replaced. The tolerance value ensures a significant distinction between all solutions stored in memory. The search, or hypothesis testing, ends when the maximum number of iterations is reached or if there is no improvement registered for solutions in *HM* over a predefined number of iterations. The final *HM* consists of a set of best-fit solutions that were found during the search.

3. Numerical experiments

Numerical experiments were conducted to test the performance of the proposed framework and solution algorithm. Four experiments are presented to illustrate different aspects of the algorithm, including the ability to (1) find specifications that closely represent observed behavior, (2) reveal additional behavioral insights from the data, which were not captured using restrictive specifications, (3) include any a priori knowledge or preference into the specification, and (4) simultaneously test many modelling hypotheses. A different dataset was used in each of the four experiments to illustrate applicability to various study contexts and data characteristics. Covering arrays were used to define hyperparameters for the experiments. The hyperparameter ranges were set based on recommendations provided in the literature (Cobos et al., 2020; Xiang et al., 2014). Similar methods have been applied in previous studies and have been found efficient for hyperparameter tuning (Ordoñez et al., 2018; Ruano-Daza et al., 2018). Table 1 presents the resulting covering arrays used in this study.

Sensitivity analysis was conducted to investigate the sensitivity of the solution algorithm to changes in hyperparameters. Fig. 3 presents the search performance using a test dataset for the twenty unique combinations of hyperparameters provided in Table 1. The convergence was seen to be achieved at BIC of 9496 for most cases, suggesting that the search is not significantly sensitive to hyperparameters for the given problem.

For the first experiment, a synthetic discrete outcome problem was designed to have full control and test the effectiveness of the proposed solution algorithm to capture underlying behavioral aspects which are known. Publicly available datasets and the corresponding specifications from the literature were used in the following three experiments to enable comparative analyses and benchmarking. For the second experiment, travel mode choice behavior was analyzed using the popular Swiss metro dataset provided by Bierlaire et al. (2001). Hyperparameters used for the first and second experiments include: $HMS = 5$; $HMCR_{min} = 0.6$; $HMCR_{max} = 0.95$; $PAR_{min} = 0.2$; $PAR_{max} = 0.8$; $\mu = 80\%$; $\Delta = 15$; and $iter_{max} = 500$.

Data from Sagebiel et al. (2018) was used in the third experiment to study the behavior of residents regarding car-free city centers in Berlin. For the fourth experiment, preferences for electricity plan was analyzed using data collected by the Electric power research Institute (Goett, 1998). The hyperparameters used for the third and fourth experiments include $HMS = 5$; $HMCR_{min} = 0.6$; $HMCR_{max} = 0.9$; $PAR_{min} = 0.3$; $PAR_{max} = 0.45$; $\mu = 75\%$; $\Delta = 30$; and $iter_{max} = 150$. Experiments were performed using 10 cores and 40 Gigabytes of Ram with GPU estimation capabilities. For estimation of random coefficients, 1,000 random draws were used for the first experiment to ensure convergence of the benchmark model. 200 draws were used for the remaining experiments. The solution algorithm was implemented in Python, and the software, along with the experiment results, can be accessed from Beeramoole et al. (2022). The lower-level objective function was estimated using searchlogit (Beeramoole et al., 2022) in-house modelling tools in Python, which is a significant extension of xlogit, developed by Arteaga et al. (2021).

4. First experiment- synthetic dataset

4.1. Data description

The Synthetic Specification used for data generation for this experiment is provided in Table 2 (column 1). A total of 10,000 discrete outcome observations from 500 unique individuals were simulated, for three alternative outcomes, and sixteen explanatory variables ($x_1, ..., x_{16}$). Variables x_1 *to* x_9 were designed to have fixed coefficients, while the remaining were provided with random coefficients that followed different distributions, as shown in column 1. To increase complexity and generate a representative case study including multiple aspects and potential real-world behavioral characteristics, non-linear transformations were applied to variables x_{12} and x_{13}. Similarly, correlation was imposed among coefficients for x_{14}, x_{15} and x_{16}. In addition, three non-significant explanatory variables x_7, x_8 and x_9, were included to test the algorithm's ability to identify variables that did not contribute to the outcomes and subsequently exclude them during the extensive hypothesis testing to facilitate the discovery of the best-fit specification. The estimated coefficients for the same specification using the maximum likelihood approach are also presented in Table 2 (column 2). This enables fair benchmarking using the specification generated with the help of the proposed solution algorithm.

4.2. Results & analysis

Table 2 (column 3) provides the experiment results. The proposed algorithm was able to identify all significant explanatory variables during the search and exclude those that were not statistically significant. For the given number of iterations, the proposed solution algorithm was able to find a specification with the goodness of fit values very close to the Benchmark Specification. The proposed algorithm was able to identify the existence of random heterogeneity in coefficients x_{12} to x_{16}. In addition, the algorithm was able to accurately identify nonlinear transformations for both x_{12} and x_{13}, along with the distributions for coefficients associated with x_{14} and x_{16}. The algorithm was also able to accurately estimate correlated coefficients for x_{14}, x_{15} and x_{16}. The main differences, however, between the Benchmark Specification and the Estimated Specification by the proposed algorithm were the coefficient distributions for other variables. The proposed solution algorithm was able to recover most behavioral aspects from the Synthetic Specification within the given constraints and search iterations. However, further research is required to ensure that misspecifications are alleviated. Adaptation of the proposed methods using analytical properties of the proposed problem to seek global optimal solutions is a promising area of open research. For example, the goodness-of-fit measures alone are not sufficient to assess if the selected random distribution adequately approximates an underlying heterogeneity in preferences. Therefore, the selection of random distributions should in fact take into consideration the properties of the corresponding density functions. This could be achieved by including distribution tests as constraints within the proposed framework, such as the semi nonparametric test proposed by Fosgerau and Bierlaire (2007) to alleviate misspecifications.

Fig. 4 shows convergence of the objective functions, BIC, and Log-likelihood (LL), over iterations for the synthetic dataset. The BICs estimated at current iteration can be observed trending towards convergence while continuing to attempt to escape potential local minima. A total of 1009 unique specifications were estimated in less than 5 h by using the GPU-assisted fast estimation techniques proposed by Arteaga et al. (2021). The BIC improved from 21,197 to 13,637 in 500 iterations, while the log-likelihood maximized from −10,585 to −6753.

5. Second experiment-travel mode choice in Switzerland

5.1. Data description

Mode choice preferences were analyzed using the stated preference dataset by Bierlaire et al. (2001) collected in Switzerland in 1998 to study the potential impact of a new transport mode – the Swissmetro. Each respondent was presented with nine hypothetical choice scenarios and was asked to choose from three transport modes (train, car, and Swiss metro). Potential explanatory variables considered for the choice analysis included[1] travel time (in minutes), travel cost (in CHF), headway for public transport modes (Train and Swiss metro), presence of luggage with traveler (no luggage, one, and more than one), seat configuration for Swiss metro (dummy variable indicating if the seats are arranged like airlines or not), dummy variable indicating if the traveler had an annual public transport ticket or not, traveler class (dummy variable to indicate first-class traveler), age, gender, income, and travel-cost bearer (self, employer, or shared by both).

5.2. Results & analysis

The objective of this experiment was to illustrate an application of the proposed extensive hypothesis testing for a transport-related analysis and to investigate its ability to reveal valuable behavioral insights relative to an available specification (Bierlaire et al., 2001). Table 3 shows the Estimated Specification by the proposed solution algorithm along with the one estimated by Bierlaire et al. (2001). The Likelihood ratio test showed a significant improvement in fit by the specification estimated using the proposed algorithm (Chi-square score = 3314; P-value <0.00001 at 95% confidence interval). Variables such as travel time, travel cost and headway were identified as important and explanatory during the search, similar to the specification by Bierlaire et al. (2001). However, significant non-linearity in the effects of travel time, cost and headway was found by the Estimated Specification. In addition, socioeconomic characteristics, and trip-related attributes, including age, presence of luggage, and availability of annual public-transport ticket were found to be significant factors influencing transport mode choices.

While the alternative-specific constants in the specification by Bierlaire et al. (2001) suggest a higher preference for the car mode, the Estimated Specification suggests a higher preference for the train mode. More than 58% of the observed sample chose train as their preferred mode. The alternative-specific constants from the Estimated Specification possibly capture the unobserved utilities for the

[1] A detailed description of the dataset can be found in Antonini et al. (2007).

Table 1
Covering arrays used for tuning hyperparameters.

Case ID	*HMS*	$HMCR_{min}$	$HMCR_{max}$	PAR_{min}	PAR_{max}	$iter_{max}$
1	5	0.5	0.99	0.2	0.7	300
2	5	0.6	0.95	0.2	0.8	500
3	5	0.8	0.95	0	0.95	400
4	5	0.9	0.97	0.15	1	400
5	10	0.5	0.85	0.25	0.7	1000
6	10	0.9	0.85	0.25	0.7	400
7	10	0.9	0.97	0.1	0.95	600
8	20	0.1	0.99	0.25	0.9	300
9	20	0.6	0.85	0.05	0.85	700
10	40	0.5	0.9	0.25	0.7	1000
11	40	0.7	0.9	0.25	1	500
12	60	0.1	0.99	0.25	0.85	1000
13	60	0.5	0.8	0.2	0.7	1000
14	60	0.6	0.8	0.1	0.9	600
15	60	0.7	0.85	0	0.8	400
16	80	0.5	0.9	0.05	0.7	1000
17	80	0.5	0.97	0	0.85	1000
18	80	0.8	0.8	0.05	0.95	2000
19	80	0.8	0.8	0.2	0.95	400
20	80	0.8	0.9	0.15	0.85	700

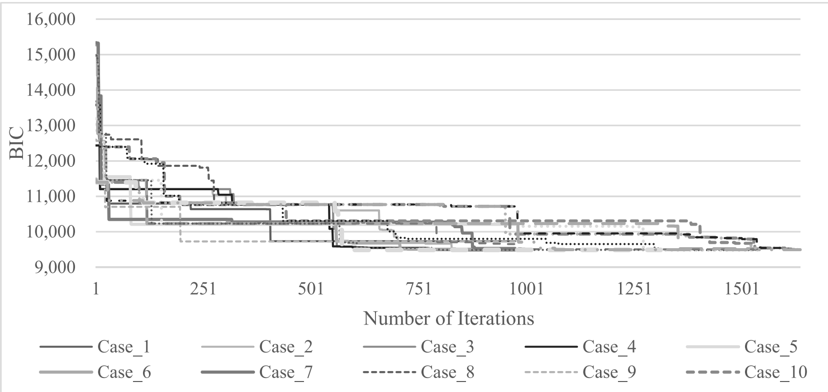

Fig. 3. Sensitivity analysis of Hyperparameters.

train due to factors such as comfort. The socioeconomic characteristics included in the Estimated Specification reveal some interesting behavioral insights. For example, the Estimated Specification suggests that as travelers age increases, they are less likely to drive, which possibly captures the age-related inconvenience associated with driving. The Estimated Specification also captured the interaction between annual public transport ticket and travel costs. On further investigation, it was found that for travelers with annual public transport ticket, the cost of the ticket was used as travel cost for analysis. As a result, the travel cost ranges were very high for travelers with annual ticket. The interaction, therefore, was able to segment the population and estimate the associated disutility of costs for those without an annual public transport ticket.

Significant heterogeneity in the effects of attributes, including travel cost, travel time and headway was captured during the extensive hypothesis testing. The random coefficients for travel time were negative for 56% of the observed sample. The random coefficients captured the associated high variances in the cost that occurred as a result of using the annual public transport ticket price as travel cost for those with the ticket. The estimated individual-specific coefficients for the other attributes suggest an overall associated disutility. However, the effect significantly varied across the observed sample. For example, the significant random coefficient for headway suggests that some travelers had a greater dislike towards waiting, indicating the likely influence of factors such as inconvenience.

To illustrate the significance of an unbiased and efficient specification, the marginal utilities from the two specifications are presented in Table 4. The mean marginal utilities estimated based on individual-specific coefficients from the specification found using the proposed solution algorithm provides an improved representation when compared to the over-estimated WTP values estimated from the restricted specification. The potential risks of misspecification can be clearly observed when restricted specifications are used without extensive hypothesis testing.

Table 5 presents a comparison of the best specification estimated by Ortelli et al. (2021) with the one estimated with the help of the proposed solution algorithm for the same dataset. The proposed extensive hypothesis testing enables to find parsimonious

Table 2
Specification used to generate the synthetic dataset along with the corresponding estimates.

		1		2			3		
		Synthetic Specification used for data generation		Benchmark Specification			Estimated Specification by the proposed extensive hypothesis testing		
Number of individuals: 500									
Number of observations: 10,000									
		Estimate	f[b]	Estimate	f[b]	t-ratio[a]	Estimate	f[b]	t-ratio[a]
Fixed Parameters									
x_1		−0.48		−0.46		−12.2***	−0.44		−11.8***
x_2		0.40		0.37		10.2***	0.37		9.9***
x_3		0.60		0.65		17.2***	0.65		17.5***
x_4		0.55		0.51		13.5***	0.49		12.7***
x_5		0.88		0.81		24.3***	0.79		24.0***
x_6		0.36		0.39		10.9***	0.37		10.5***
x_7		0.00		0.00					
x_8		0.00		0.00					
x_9		0.00		0.00					
Random Parameters									
x_{10}	mean	0.53		0.52		13.8***	0.49		13.8***
	s.d.	0.31	u	0.45	u	3.2**			
x_{11}	mean	0.90		0.89		21.3***	0.85		22.1***
	s.d.	0.32	t	0.92	t	5.8***			
Random Parameters with nonlinear transformations									
x_{12}	mean	−1.49		−1.30		−22.8***	−0.57		−20.2***
λ_{12}		0.4		0.4					
	s.d.	1.37	n	1.08	n	22.1***	1.21	t	22.0***
x_{13}	mean	0.61		0.55		9.86***	0.56		9.4***
λ_{13}		0.3		0.3			0.3		
	s.d.	1.29	n	1.09	n	21.1***	2.56	t	22.3***
Random Correlated Parameters									
x_{14}		0.77	n	0.65	n	12.5***	0.64	n	12.0***
x_{15}		0.99	n	0.99	n	18.1***	0.95	t	16.1***
x_{16}		−1.27	n	−1.19	n	−20.5***	−1.15	n	−17.9***
Cholesky factors									
x_{14}, x_{14}		1.00		1.00		18.7***	0.97		17.8***
x_{15}, x_{14}		0.25		0.40		5.6***	0.40		5.9***
x_{15}, x_{15}		0.97		0.93		15.9***	2.21		16.7***
x_{16}, x_{14}		0.40		0.41		5.1***	0.42		5.3***
x_{16}, x_{15}		0.413		0.38		4.4***	0.91		4.4***
x_{16}, x_{16}		0.82		0.87		13.5***	0.89		13.8***
LL				−6701			−6753		
BIC				13,545			13,637		

[a] weakly significant ($p < 0.10$, $t > 1.645$), ** = significant ($p < 0.05$, $t > 1.96$), *** = strongly significant ($p < 0.01$, $t > 2.58$).
[b] n = normal; u = uniform; t = triangular; ln = lognormal.

specifications with improved explanatory power. Further, the out-of-sample measures also show a significant improvement, in terms of model interpretability, over the approach proposed by Ortelli et al. (2021) to assist in the specification search.

Evidently, the Estimated Specification provides substantial information on travel behavior compared to a the specification reported in literature. While goodness-of-fit is a crucial measure that indicates whether a given specification is representative of the observed behavior, the best-found model may not necessarily capture the behavioral aspects essential for a given study context. Therefore, the proposed solution algorithm is designed to provide multiple solutions at the end of the search, with varying properties and goodness-of-fit. Table 6 provides three such selected specifications found during the extensive hypothesis testing. The analyst can utilize these solutions as starting points and continue the model development process in line with the study context. As illustrated in the experiment, the proposed solution algorithm enables extensive hypothesis testing to provide detailed insights into behavior, which could be useful in developing new policy alternatives for transport infrastructure planning.

The final model was selected after a strategic hypothesis testing from a total of 1375 specifications, which were estimated in 12 h. Fig. 5 shows convergence of the objective functions, BIC from 10,624 to 9,913, and LL from −5264 to −4905.

6. Third experiment-preferences for a car-free city center

6.1. Data description

A stated preferences choice survey was conducted in Berlin to study the behavior of residents regarding car-free city centers (Gundlach et al., 2018). A total of 347 respondents were asked to choose between three alternatives, wherein two were for car-free city

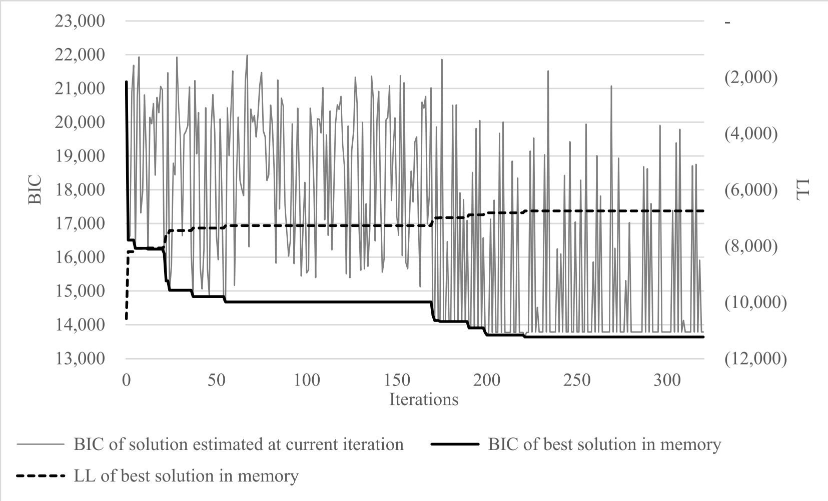

Fig. 4. BIC versus Iterations for the experiment using the synthetic dataset.

Table 3
Specifications found by Bierlaire et al. (2001) and by the proposed extensive hypothesis testing.

		Specification by Bierlaire et al. (2001)		Estimated Specification by the proposed solution algorithm			
Number of respondents: 924							
Number of observations: 8316							
Parameter		Estimate	t-ratio[a]	λ_k[c]	Estimate	t-ratio[a]	f[b]
For Swiss metro							
Seat configuration		0.16	2*				
Travel cost	mean	−0.001	−19.6***	Log	−3.01	−15.1***	
	s.d.				7.98	22.8***	*u*
Travel time	mean	−0.01	−24.3***	Square root	−0.58	−23.3***	
	s.d.				1.00	22.0***	*t*
Headway	mean	−0.01	−7.8***	Log	−0.59	−10.0***	
	s.d.				0.58	9.1***	*n*
For Car							
Mode-specific constant		0.062	1.2		−3.02	−6.6***	
Age	mean				−0.61	−2.5*	*ln*
	s.d.				0.86	8.73***	
Luggage	mean	−0.12	−2.5**		−2.84	−7.8***	
	s.d.				5.37	14.2***	*n*
Travel cost	Mean	−0.001	−19.6***	Log	−3.01	−15.1***	
	s.d.				7.98	22.8***	*u*
Travel time	mean	−0.01	−24.3***	Square root	−0.58	−23.3***	
	s.d.				1.00	22.0***	*t*
For Train							
Mode-specific constant		−1.16	−10.4***		2.44	14.3***	
Annual public transport ticket		7.49	21.9***				
Age		0.19	6.1***				
Travel cost for travelers without annual public transport ticket	Mean				−0.06	−13.0***	
	s.d.				0.07	12.1***	*u*
Travel cost	Mean	−0.001	−15.8***	Log[c]	−3.01	−15.1***	
	s.d.				7.98	22.8***	*u*
Travel time	Mean	−0.01	−15***	Square root[c]	−0.58	−23.3***	
	s.d.				1.00	22.0***	*t*
Headway	Mean	−0.007	−7.8***	Log[c]	−0.59	−10.0***	
	s.d.				0.58	9.1***	*n*
LL		**−6565**				**−4908**	
BIC		**13,211**				**9913**	

[a] * = weakly significant ($p < 0.10$, $t > 1.645$), ** = significant ($p < 0.05$, $t > 1.96$), *** = strongly significant ($p < 0.01$, $t > 2.58$).
[b] n = normal; u = uniform.
[c] Travel cost, travel time and headway entered as linear-in-parameters in specification by Bierlaire et al. (2001), whereas non-linearly in the Estimated Specification by the proposed solution algorithm.

centers with varying attribute levels, and the third alternative was the status quo (or “as today”). A panel dataset was collected by presenting each respondent with nine scenarios, which varied based on: (i) availability of biking infrastructure (as today, availability of bike lanes, separate car-free road network for cyclists), (ii) walking distance to the nearest public transport stop (as today, 6 min, 3 min), (iii) public transport frequency (as today, higher than today, highest), (iv) park & ride facilities at destinations close to city-center (as today, unguarded parking facility, guarded parking lots), (v) price of public transport (free, 75% lesser than today, 50% less, 25% less, as today, 25% more), and (vi) recreational areas in the city center (as today, 20% more, 40% more). Socio-economic characteristics included car ownership, resident within the city-center, frequency of public transport usage, and the respondent’s gender and age. A car-free factor (CF = 1), along with the socio-economic characteristics were included in the utility specified for the car-free city center alternative.[2]

6.2. Results & analysis

This experiment was performed to illustrate the flexibility of the proposed extensive hypothesis testing algorithm to allow the use of a priori knowledge in model development. Some of the assumptions imposed on Gundlach’s Specification were included as pre-specifications in the experiment. The car-free factor was forced into the specification, and socio-economic variables (interacted with CF) were restricted to be estimated with fixed coefficients. Many characteristics in the dataset, such as income, bicycle ownership and number of children, which were not previously included in Gundlach’s Specification, were also considered during the model search. Table 7 presents the Estimated Specification found by the proposed algorithm along with Gundlach’s Specification for comparison.

A parsimonious specification, Estimated Specification, with a better fit was found with the help of the proposed algorithm, with several variables from Gundlach’s Specification eliminated after extensive and rigorous hypothesis testing. The likelihood-ratio test values suggest that the specification estimated by the proposed algorithm achieved a significant improvement in fit (Chi-square score = 370.72; P-value <0.001 at a 95% confidence level). Among the socio-economic characteristics considered during the search, only the car-ownership factor was found to be statistically significant and therefore retained in the final model. Using the pre-specification constraints enabled by the proposed solution algorithm, the analyst can still retain the excluded socio-economic characteristics if they are essential factors for policy analysis. Further, other specifications in memory, as explained in section 5.2, can also be considered for analysis if they provide greater relevance to the study.

For variables estimated with random coefficients, the mean values from both Estimated and Gundlach’s Specifications had the same signs. However, considering that they were estimated based on different distributional assumptions, their behavioral interpretation varied considerably for some variables. The random coefficients from Gundlach et al. (2018) implied that a significant portion of the sample associated a disutility with cycling infrastructure (33% for availability of bike-lanes, and 39% for separate bike-lanes) for choosing a car-free city center. In contrast, the Estimated Specification showed that all of the observed samples associated a utility with the availability of bike lanes, and only 23% of the sample disliked having a separate car-free cycling infrastructure. Considering that the study area is Berlin, which is well-known for its large bicycle-loving society, the results from the Estimated Specification were more consistent with observed behavior. Further, given that only 19% of the sample owned cars, the higher preference for car-free cycling infrastructure in the city center is justified. The heterogeneity in taste for the availability of recreational facilities is similarly captured by both models. The public transport fare coefficient in Gundlach et al. (2018) was estimated with a lognormal distribution to restrict signs. However, even without imposing such distributional assumptions, the corresponding random coefficient in the Estimated Specification indicated that all of the observed samples associated a disutility with public transport cost. The Estimated Specification also indicates a significant variation in the public-transport fare preferences, which potentially captures the influence of individual-specific public-transport usage patterns.

While the specification by Gundlach et al. (2018) found a significant disutility for a 6-min walking distance to the nearest public transport stop, the Estimated Specification found a significant utility for a 3-min walking distance. As discussed by Gundlach et al. (2018), 84% of the observed sample reported their current walking distance, from residence to the nearest public transport stop, as less than 6 min. Therefore, the preference for a car-free center would evidently increase if this distance further reduced. The Estimated Specification found that the availability of park & ride facilities or improvement in public transport frequency did not significantly influence the preferences for car-free city centers and therefore were excluded from the final model.

In addition, allowing for correlated coefficients was found to significantly improve model fit. Table 8 presents the correlation matrix from the Estimated Specification. Critical insights for policy design were obtained from the correlation analysis. A counter-intuitive finding is the strong correlation found between preferences for shorter walking distances to public transport and availability of recreational areas. The correlation potentially captures the latent preferences of residents for a greener and sustainable city center.

Fig. 6 presents a comparative analysis of the estimated choice probabilities for different policy alternatives as described in Gundlach et al. (2018). The specification estimated using the proposed extensive hypothesis testing framework provided a similar preference for car-free city centers, to the specification by Gundlach et al. (2018). A higher sensitivity was observed towards public transport (PT) fares, which is indicated by the lowest estimated choice probabilities for a policy alternative with strong measures but increased PT fare. Policy alternatives that focus on improving the bicycle network is preferred the most. However, a significant

[2] Refer Gundlach et al. (2018) for a detailed description of the data and the collection methods adopted. The data is open source and can be found in Sagebiel et al. (2018).

Table 4
Mean marginal utilities estimated from Bierlaire et al. (2001) and the Estimated Specification.

	Mean willingness to pay additional travel costs in CHF	
	Estimates using Bierlaire et al. (2001)'s specification	Estimates using the specification by the proposed extensive hypothesis testing
To reduce travel time by 1 min	10	0.59
To reduce headway by 1 min	10	0.005

Table 5
Comparison of Specifications estimated by Ortelli et al. (2021) and the proposed extensive hypothesis testing.

	LL	Model parameters	Out-of-sample LL[a]
Specification by Ortelli et al. (2021)	−6136	28	−1568
Specification estimated by the proposed solution algorithm	−4905	14	−1239

[a] The out-of-sample LL was estimated using a testing dataset which included 2079 observations that were not included in model training.

Table 6
Selected model specifications found during the extensive hypothesis testing along with their best goodness-of-fit.

BIC	Variable	Coefficient type	Transformation	Random coefficients
9925	Constant	Alternative-specific		
	Travel cost for travelers without annual public-transport ticket	Alternative-specific		uniform
	Gender	Alternative-specific		
	Age	Alternative-specific		logarithm
	Presence of luggage			normal
	Travel time	Generic	Logarithmic	triangular
	Travel Cost	Generic	Square root	uniform
	Headway	Generic	Logarithmic	normal
9945	Constant	Alternative-specific		
	Seats	Alternative-specific		
	Presence of luggage	Alternative-specific		normal
	Travel cost for travelers without annual public-transport ticket	Alternative-specific		uniform
	Travel time	Generic	Logarithmic	triangular
	Travel Cost	Generic	Square root	uniform
	Headway	Generic	Logarithmic	normal
9983	Constant	Alternative-specific		
	Seats	Alternative-specific		
	Income	Alternative-specific		normal
	Travel cost for travelers without annual public-transport ticket	Alternative-specific		triangular
	Presence of luggage	Alternative-specific		triangular
	Travel time	Generic	Logarithmic	uniform
	Travel Cost	Generic	Logarithmic	triangular
	Headway	Generic		triangular

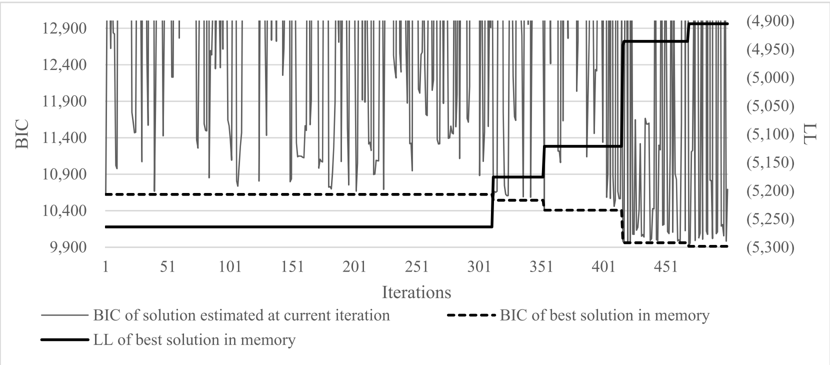

Fig. 5. BIC vs Iterations for the experiment regarding the mode-choice preferences using Swiss metro dataset.

difference can be observed in the estimated choice probabilities from the two models for the "status quo" policy alternative. The improved specification suggests that the observed sample has higher preference towards car-free city centers even in status-quo conditions, wherein there is no other improvement introduced. While the two models provide very similar sensitivity patterns for other policy alternatives, the analyst efforts required to estimate the improved specification was significantly less than using conventional methods. Further, pre-specifications can also be used in similar manner, as shown in this experiment, to test other policy measures by fixing specific aspects of the model that are essential for analysis.

The total estimation time to find the Estimated Specification was 2 h and 21 min, in which 20 potentially explanatory variables were included during extensive hypothesis testing, investigating the presence of taste heterogeneity and correlation while considering knowledge provided by the analyst. Fig. 7 illustrates the convergence of the objective functions over iterations. A total of 904 specifications were estimated with a 7% improvement in BIC and 7.5% improvement in the Log-likelihood. For every x number of variables in the data, there exists 2^x number of unique combinations. An exhaustive search is not performed in practice. However, if an exhaustive hypothesis testing was performed to determine potential explanatory variables from 20 considered, the total unique specifications would easily go up to $2^{20} = 1,048,575$. These unique combinations would further increase if the testing included non-linear transformations, random coefficients, and potential correlation. In practice, the specification search is often restricted based on knowledge, experience, and sometimes subjective selection of hypothesis (Ortelli et al., 2021; Paz et al., 2019), while still requiring the analyst to spend substantial time testing specifications.

7. Fourth experiment-electricity-supplier choice in California

7.1. Data description

For the fourth experiment, the choice of electricity supplier was analyzed using data collected in California by the Electric Power Research Institute (Goett, 1998). A stated-preference survey was conducted on 361 residential customers to study their preferences regarding electricity plans. The panel dataset includes a total of 4308 observations wherein each customer faced up to 12 choice scenarios with four different plans to choose from. Each choice scenario was designed using six attributes, including a fixed price for an electricity plan (7 or 9 cents/kWh), contract length during which a penalty is imposed if the customer chooses to switch plans (no contract, 1 year or 5 years), a dummy variable indicating if the supplier was well-known, time of the day rates (11 cents/kWh from 8AM to 8PM and 5 cents/kWh from 8PM to 8AM), seasonal rates (10 cents/kWh for summer, 8 cents/kWh for winter and 6 cents/kWh in spring and fall) and, a dummy variable indicating if the supplier was a local.

7.2. Results & analysis

The proposed extensive hypothesis testing framework was utilized to generate and estimate a discrete choice model while simultaneously investigating the presence of unobserved heterogeneity in preferences, potential correlation between variables and their non-linear interactions. Table 9 shows the best model, Estimated Specification, found by the proposed solution algorithm along with the specifications found by Revelt and Train (2000) and Paz et al. (2019).

The BIC and Likelihood values show significant improvement in fit for the Estimated Specification relative to the one by Revelt and Train (2000) as well as Paz et al. (2019). The specification found in this study performs significantly better than those in the literature, particularly due to the extensive hypothesis testing, which includes investigating non-linear and correlated effects of attributes on choice behavior. The likelihood ratio test (Chi-square score = 418; p-value <0.001) suggests that the compared specifications are significantly different. However, the solution algorithm identified significant explanatory variables as same as those in the specification by Revelt and Train (2000). The associated disutility of fixed price of an electricity plan, seasonal and time of the day rates, and contract length are accurately captured by the negative coefficients and are consistent with expectations based on general economic theory. In addition, the model generated with the help of the proposed framework includes non-linear effects of fixed price on the choice of plans. Fig. 8 suggests that the linear specification by Paz et al. (2019) overestimates the disutility as the fixed price increases. On further investigation, it was observed that plans offered with lesser fixed prices entailed a variable price component that includes time-of-the day and seasonal rates. However, the variable costs were 0 for the plans with higher fixed prices. Therefore, it is likely for customers to perceive a relatively lesser disutility at higher fixed prices due to the associated concessions in the variable price component.

The proposed algorithm tested and found significant heterogeneity for all variables, including price. Various distributions were tested for coefficients, including normal, truncated-normal, lognormal, uniform, and triangular, and those that provided the best-fit were retained. The random parameters in the Estimated Specification show that almost all customers associated a disutility with cost-related attributes, which aligns with the interpretation from Paz et al. (2019). In addition, the heterogenous preferences for time-of-the-day and seasonal rates reveal how customers' electricity-usage patterns affect their preferences. The coefficients in the Estimated Specification also show that the observed sample prefers local over known suppliers but with a significant variation in taste. The heterogeneity possibly reflects the influence of previous experience with local and well-known suppliers. The random coefficients also show that none of the samples prefer unknown suppliers indicating that the entrant offers were not attractive enough for customers to prefer new suppliers, which aligns with the inferences derived from Revelt and Train (2000).

The Estimated Specification shows that the observed sample associated a disutility with contract length. Further, the estimated variances for the contract length capture varying levels in tastes of customers. While a significant portion of the sample perceived a

Table 7
Specifications found by Gundlach et al. (2018) and by the proposed extensive hypothesis testing.

		Gundlach's Specification[a]			Estimated Specification by the proposed extensive hypothesis testing		
Number of respondents: 347							
Number of observations: 3123							
Parameter		Estimate	t-ratio[c]	f[b]	Estimate	t-ratio[c]	f[b]
Socio-economic characteristics							
CF		3.61	12.3***	–	0.52	3.2***	
CF × owns car		−3.11	−30.6***	–	−2.04	−7.2***	
CF × public transport usage		−0.47	−10.9***	–			
CF × male		−0.50	−6.8***	–			
CF × resident		0.17	2.3**	–			
CF × age		0.001	−0.1	–			
Attributes of city center							
Recreational areas	Mean	0.38	9.9***		0.45	6.4***	
	s.d.	0.89	32.6***	*n*	0.90	24.6***	*n*
Public Transport fare	mean	−0.43	−36.2***		−0.73	−14.2***	
	s.d.	2.02	336***	*n*	0.69	22.8***	*u*
Bike lanes available	mean	1.39	16***		1.49	14.6***	
	s.d.	3.12	34.3***	*n*			-
Separate car-free bike lanes	mean	1.7	17.8***		1.8	10.1***	
	s.d.	5.98	68***	*n*	2.49	23.2***	*n*
Three-min walking to nearest public transit	Mean	0.55	6.6***		0.64	6***	
	s.d.	1.49	16***	*n*	0.8	28***	*u*
Six-min walking to nearest public transit	Mean	−0.14	−1.5				
	s.d.	1.65	13.6***	*n*			
Unguarded parking facility for Park & Ride	mean	−0.06	−0.7				
	s.d.	2.07	18.4***	*n*			
Guarded parking facility for Park & Ride	mean	−0.40	−4.6***				
	s.d.	1.58	16.1***	*n*			
Highest frequency of public transport	mean	0.26	3.1***				
	s.d.	1.60	14.9***	*n*			
Higher frequency of public transport	mean	0.22	2.6***				
	s.d.	2.79	29.3***	*n*			
LL		**−2137**				**−1977**	
BIC		**4366**				**4053**	

[a] * = weakly significant ($p < 0.10$, $t > 1.645$), ** = significant ($p < 0.05$, $t > 1.96$), *** = strongly significant ($p < 0.01$, $t > 2.58$).
[b] n = normal; u = uniform.
[c] The specification developed in Gundlach et al. (2018) is used for a comparative analysis. The estimated coefficients, however, are slightly different from that reported in literature due to the difference in the number of observations used for model estimations.

Table 8
Correlation matrix for Berlin using the Estimated Specification by the proposed algorithm.

	Separate car-free bike lanes	Three-min walking to nearest public transit	Public Transport fare	Recreational areas
Separate car-free bike lanes	1.0			
Three-min walking to nearest public transit	0.01	1.0		
Public Transport fare	−0.3	−0.5	1.0	
Recreational areas	0.28	**0.8**	−0.5	1.0

Note: Strong correlations highlighted in bold.

total disutility, the remaining sample perceived a utility from contract length due to the influence of insurance. Plans offered with contracts had an insurance that locked the fixed price for the stipulated period, which is likely to make contracts attractive for some customers. However, the model in literature estimated a normally distributed coefficient to capture the effect of contract length, whereas the proposed extensive hypothesis testing estimated the corresponding coefficient using a triangular distribution. The probability density plots for the two coefficients are presented in Fig. 9 for comparison.

The mean and standard deviation estimates for both normal and triangular distributions were close to −0.2 and 0.4, respectively. However, the two distributions indicate a substantial difference in the proportion of sample that perceives utility from contract length. While the normal distribution indicates that 29% of the sample perceived a utility from contract length, the triangular distribution suggests that only 10% of the sample are likely to find contract length attractive. Consequently, the two specifications can provide substantially different market shares and associated behavioral inferences for different policy scenarios. These findings further validate the need for an extensive hypothesis testing framework to improve model specification and associated behavioral outcomes.

Table 10 presents a correlation matrix using the Estimated Specification, which provided additional insights about customer

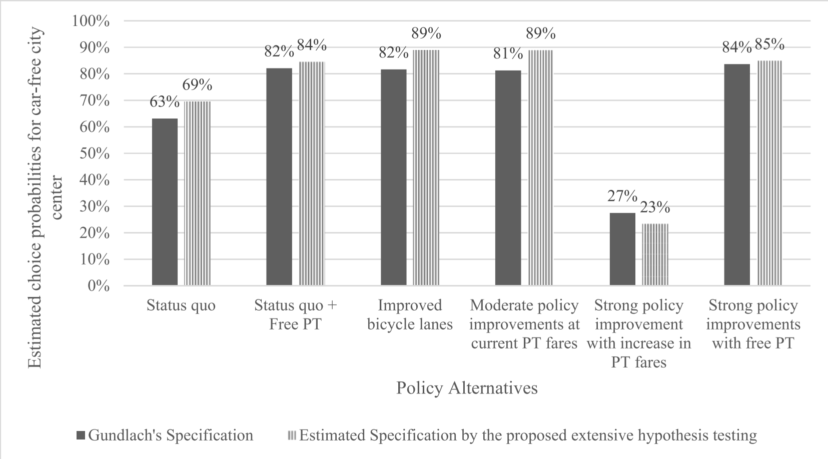

Fig. 6. Estimated choice probabilities for different policy alternatives of car-free city centers.

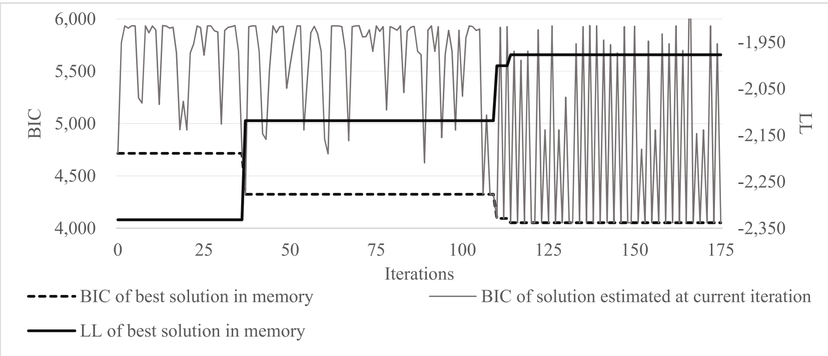

Fig. 7. BIC versus iterations for the experiment regarding car-free center preferences in Berlin.

Table 9
Specifications found by Revelt and Train (2000), Paz et al. (2019) and the proposed extensive hypothesis testing.

		Specification by Revelt and Train (2000)			Specification by Paz et al. (2019)			Estimated Specification by the proposed extensive hypothesis testing			
Number of respondents: 361											
Number of observations: 4308											
Parameter		Estimate	t-ratio[a]	f[b]	Estimate	t-ratio[a]	f[b]	λ_k[c]	Estimate	t-ratio[a]	f[b]
Fixed price	mean	−0.90	−27.4***		−0.96	−27.3***		log	−7.02	−22.5***	
	s.d.				0.18	16.1***	*n*		4.92	21***	*u*
Length of contract	mean	−0.21	−10***		−0.23	−16.2***			−0.20	−8.9***	
	s.d.	0.4	19.1***	*n*	0.38	20.3***	*n*		0.40	22.5***	*t*
Time of day rates	mean	−8.74	−19.6***		−9.35	−30.4***			−16.00	−25.1***	
	s.d.	2.57	−28.7***	*n*	2.26	17.9***	*n*		10.89	20.8***	*u*
Seasonal rates	mean	−9.05	−29.1***		−9.33	−30.7***			−16.16	−25***	
	s.d.	2.014	17.4***	*n*	1.53	11.3***	*n*		10.53	20.8***	*u*
Local supplier	Mean	2.16	20.7***		2.30	26.1***			2.43	17.7***	
	s.d.	1.61	14.9***	*n*	2.80	18.4***	*u*		2.28	24.5***	*tn*
Well-known supplier	Mean	1.55	18.4***		1.56	23.2***			1.92	19.2***	
	s.d.	1.05	12.1***	*n*	1.1	14.9***	*n*		1.4	22.3***	*t*
LL		**−3938**			**−3914**				**−3705**		
BIC		**7942**			**7928**				**7568**		

[a] * = weakly significant ($p < 0.10$, $t > 1.645$), ** = significant ($p < 0.05$, $t > 1.96$), *** = strongly significant ($p < 0.01$, $t > 2.58$).
[b] Tn = truncated- normal; u = uniform; t = triangular.
[c] Fixed price entered as linear-in-parameters in the specification by Paz et al. (2019), whereas non-linearly (log) in the Estimated Specification by the proposed solution algorithm.

preferences in comparison to the specification found by Paz et al. (2019). A strong correlation can be observed between most of the variables indicating that the assumption of uncorrelated coefficients in the Specification by Paz et al. (2019) is violated. A positive correlation between cost-related variables can be observed, suggesting that the associated disutility increases with an increase in usage. In addition, the customers' utility for local suppliers increases with the utility for well-known suppliers. The associated disutility from cost-related variables decreases as the utility from local and well-known suppliers increase, suggesting that customers are willing to negotiate on the cost if suppliers are local and well-known. Insights from correlation, which can be helpful in designing more attractive plans, can be easily missed when restricted specifications are used without extensive hypothesis testing.

The mean marginal utilities obtained using the Estimated Specification are compared with the analysis conducted by Revelt and Train (2000) in Table 11 to illustrate the significance of the proposed framework. The mean WTP values are estimated from the individual-specific values, and therefore representative of the observed sample. Table 3 suggests a significant difference between WTP estimates from both models. The higher sensitivity of customers for time of day and seasonal rates is captured by both models, but with significantly lower values given by the Estimated Specification.

Fig. 10 shows the convergence of the objective functions over iterations for the best solution found by the proposed algorithm for electricity choice behavior. The BIC was minimized from 9753 to 7568 while Log-likelihood improved from −4859 to −3705 in 150 iterations. The total execution time was 6 h, 832 models were estimated, and the best five were saved in the memory.

8. Conclusion

This study proposed an extensive hypothesis testing framework to assist analysts during the estimation of discrete outcome models, focusing on mixed-Logit specifications. The framework includes a mathematical programming formulation and a bi-level constrained optimization algorithm involving a maximum likelihood estimation and a population-based metaheuristic to simultaneously consider multiple modelling decisions. This study contributes to the literature by proposing a framework that enables extensive hypothesis testing to simultaneously consider potential explanatory variables, their functional forms, the distributional assumptions of coefficients, and correlations. The proposed formulation and solution algorithm provides flexibility to pre-specify or impose certain modelling aspects to enable testing of specific hypotheses or ensure compliance with well-established theories from relevant fields, including economics and behavioral sciences. BIC was chosen as an upper-level objective function to seek parsimonious specifications while addressing potential overfitting (Khadka and Paz, 2017; Paz et al., 2019; Schwarz, 1978). The proposed framework is not envisioned to substitute or replace the analyst, but rather to enable extensive hypothesis testing considering simultaneously non-linear effects, unobserved heterogeneity, and correlations. Future research is recommended to consider alternative objective functions to seek model generalizability and/or out-of-sample estimates. Similarly, future extensions are recommended to consider simultaneously multiple objectives including but not limited to in- and out-of-sample estimates to address overfitting, maximize interpretability, and extract as much insights as possible from the data.

Four experiments, including different data sets and behavioral processes, were conducted to illustrate the significance of the proposed solution algorithm in discovering important insights that could be critical for behavioral analysis. All experiments include benchmark models either developed by independent research teams and published in the literature or created using synthetic data to enable full knowledge and control. The results illustrate the significance of the extensive hypothesis testing for discovering important influential or contributory factors, along with hidden patterns of nonlinearity, heterogeneity, and correlation, which can potentially be overlooked due to limited or biased searches. All experiments showed the ability of the proposed algorithm to generate specifications that closely capture empirical behavior within a reasonable time.

A primary goal of any modelling is to capture as much information, insights, empirical truth, and underlying behavior as possible. Results using the synthetic dataset suggests that the proposed algorithm can capture most of the behavioral information from the data, which otherwise would require an exhaustive search. However, further research is required to capture all aspects of the behavior available in the data. Perhaps, the integration of metaheuristics with deterministic methods as expansions to the one proposed in this study is required to capture the complete behavioral characteristics. Similarly, a single "best" specification that explains all aspects of an empirical dataset may not exist. Hence, as shown by the results using the Swiss metro dataset, the proposed solution algorithm can generate multiple acceptable solutions with varying properties and goodness-of-fit. For example, one specification captured those factors as significant and explanatory, which were excluded from the specification used in the literature, whereas others captured the effect of unobserved factors on mode choice preferences using different non-linear transformations or coefficient distributions. The results validate that the proposed framework can act as a decision-support tool during model development by providing relevant and meaningful starting points to the analyst. However, a thorough knowledge of the problem context and a detailed investigation of the data is still essential to support the final model selection and interpretation. This knowledge can be included both using constraints available in the framework as well as through evaluation and conventional modelling processes considering insights and outputs obtained from the proposed solution algorithm.

Regardless of the model development approach, analyst's knowledge and experience are necessary to guide the specification search in line with the study context. Results using the Berlin dataset on car-free city center preferences illustrate the flexibility of the proposed framework to utilize knowledge from the analyst and seek additional insights from the data by conducting constrained hypothesis testing. For example, the proposed semi-supervised hypothesis testing was able to find a strong correlation between the preferences of Berlin residents for improved public-transport accessibility and recreational facilities, which are important insights for policy analysis. The final specification found using the proposed algorithm for electricity choices in California also illustrate the significance of extensive testing to obtain more and better insights. An extension to improve the proposed framework is the inclusion of advanced mixed logit specifications including flexible error structures to capture unobserved heterogeneity, and to investigate the

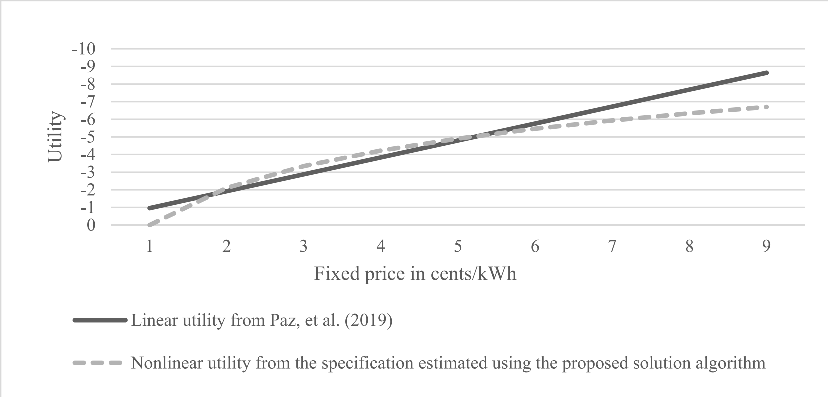

Fig. 8. Difference in the change in utility for fixed price between the Estimated Specification and the one by (Paz et al., 2019).

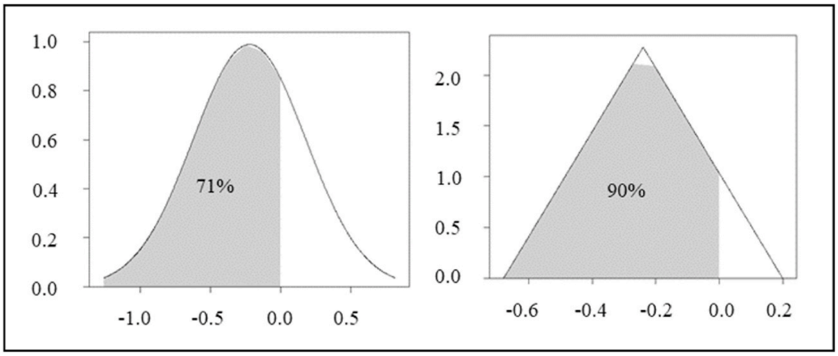

Fig. 9. Normal and triangular coefficient distributions for the effect of contract length on electricity supplier choice.

Table 10
Correlation matrix using the Estimated Specification for electricity choice behavior in California.

	Contract length	Local supplier	Fixed price	Seasonal rates	Time of day rates	Well-known supplier
Contract length	1					
Local supplier	0.43	1.00				
Fixed price	0.19	**0.95**	1.00			
Seasonal rates	0.18	**0.91**	**0.98**	1.00		
Time of day rates	0.20	**0.91**	**0.96**	**0.97**	1.00	
Well-known supplier	0.26	**0.77**	**0.84**	**0.79**	**0.79**	1

Note: Strong correlations highlighted in bold.

Table 11
Mean marginal utilities estimated from Revelt and Train (2000) and the Estimated Specification.

	Mean willingness to pay additional fixed price in cents per kWh	
	Estimates using Revelt & Train's specification	Estimates using the specification by the proposed extensive hypothesis testing
To reduce contract length by 1 year	0.23	0.12
To reduce time of day rates by 1 cent/kWh	9.74	1.50
To reduce seasonal rates by 1 cent/kWh	10	1.28
To choose plan from a local Supplier	2.42	0.02
To choose plan from a well-known supplier	1.74	0.25

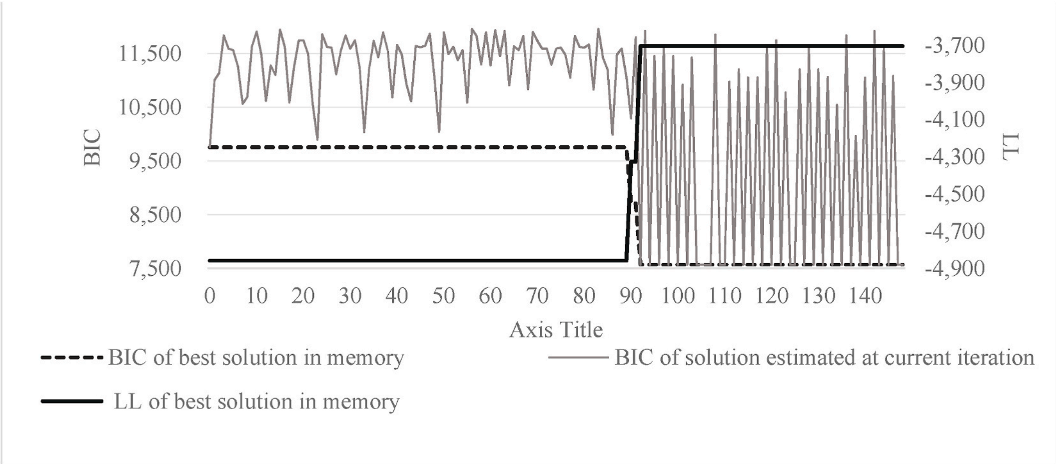

Fig. 10. BIC versus Iterations for the experiment regarding electricity choice behavior.

effect of latent preferences on observed behavior (Ben-Akiva et al., 2002). Inclusion of advanced specifications is likely to significantly increase computational complexity given the large number of model parameters involved. Hence, this extension is proposed for future research.

Author statement

Prithvi Bhat Beeramoole: Study conception, Methodology, Software implementation, Article preparation. Cristian Arteaga: Study conception, Methodology, Software implementation. Alban Pinz: Article preparation and intellectual inputs to improve the content. Md. Mazharul Haque: Article preparation and intellectual inputs to improve the content. Alexander Paz: Study conception, Methodology, Article preparation.

Declaration of competing interest

The authors have no conflicts of interest.

Data availability

Data will be made available on request.

Acknowledgements

This study was conducted as part of the "Transport Academic Partnership" between the Queensland Department of Transport and Main Roads and the Queensland University of Technology. The authors acknowledge the support and resources offered from the partnership. The authors also acknowledge Mr. Ryan Kelly for his help with the implementation of the discrete choice modelling package in Python. The authors are in process of implementing an open-source Python package for the assisted estimation of discrete choice models, along with documentation and usage examples (https://pypi.org/project/searchlogit/).

References

Akaike, H., 1998. Information Theory and an Extension of the Maximum Likelihood Principle. In: Parzen, E., Tanabe, K., Kitagawa, G. (Eds.), Selected Papers of Hirotugu Akaike. Springer Series in Statistics. Springer, New York, NY. https://doi.org/10.1007/978-1-4612-1694-0_15.

Alia, O.M., Mandava, R., 2011. The variants of the harmony search algorithm: an overview. Artif. Intell. Rev. 36 (1), 49–68. https://doi.org/10.1007/s10462-010-9201-y.

Anowar, S., Eluru, N., 2018. Univariate or multivariate analysis for better prediction accuracy? A case study of heterogeneity in vehicle ownership. [Article]. Transportmetrica A-Transport Science 14 (8), 635–668. https://doi.org/10.1080/23249935.2017.1422045.

Antonini, G., Gioia, C., Frejinger, E., 2007. Swissmetro: Description of the Data. Retrieved from. https://transp-or.epfl.ch/pythonbiogeme/examples/swissmetro/swissmetro.pdf.

Arouna, A., Dabbert, S., 2012. Estimating rural households' willingness to pay for water supply improvements: a Benin case study using a semi-nonparametric bivariate probit approach. Water Int. 37 (3), 293–304. https://doi.org/10.1080/02508060.2012.687507.

Arteaga, C., Park, J., Beeramoole, P., Paz, A., 2021. Xlogit: an Open-Source Python Package for GPU-Accelerated Estimation of Mixed Logit Models. Under review. Retrieved from. https://github.com/arteagac/xlogit.

Beeramoole, P., Kelly, R., Arteaga, C., Paz, A., 2022. Searchlogit [Computer software]. Retrieved from. https://pypi.org/project/searchlogit/.

Beeramoole, P.B., Arteaga, C., Pinz, A., Haque, M.M., Paz, A., 2022. Software and Experiment Results: Extended Hypothesis Testing during the Estimation of Mixed Logit Models [Computer Software]. Retrieved from. https://github.com/PrithviBhatB/Extended-hypothesis-testing-during-the-estimation-of-mixed-Logit-models-.

Ben-Akiva, M., Daly, A., Gunn, H., 1987. Destination choice models: design and appraisal. In: Proceedings of Seminar Transportation Planning Methods: C Held at the PTRC Summer Annual Meeting, P290, 7-11 September.

Ben-Akiva, M., Mcfadden, D., Train, K., Walker, J., Bhat, C., Bierlaire, M., Munizaga, M.A., 2002. Hybrid choice models: progress and challenges. Market. Lett. 13 (3), 163–175. https://doi.org/10.1023/A:1020254301302.

Ben-Akiva, M.E., Lerman, S.R., Lerman, S.R., 1985. Discrete Choice Analysis: Theory and Application to Travel Demand, 9. MIT press.

Bierlaire, M., 1998. Discrete choice models. In: Operations Research and Decision Aid Methodologies in Traffic and Transportation Management. Springer, pp. 203–227.

Bierlaire, M., Axhausen, K., Abay, G., 2001. The Acceptance of Modal Innovation: the Case of Swissmetro. Retrieved from. https://www.researchgate.net/publication/37456549_The_acceptance_of_modal_innovation_The_case_of_Swissmetro.

Blaine, T.W., Lichtkoppler, F.R., Bader, T.J., Hartman, T.J., Lucente, J.E., 2015. An examination of sources of sensitivity of consumer surplus estimates in travel cost models. J. Environ. Manag. 151, 427–436. https://doi.org/10.1016/j.jenvman.2014.12.033.

Box, G.E.P., Cox, D.R., 1964. An analysis of transformations. J. Roy. Stat. Soc. B 26 (2), 211–243. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x.

Brathwaite, T., Vij, A., Walker, J.L., 2017. Machine Learning Meets Microeconomics: the Case of Decision Trees and Discrete Choice. *arXiv preprint*. Retrieved from. https://arxiv.org/abs/1711.04826.

Ceylan, H., Ceylan, H., 2012. A Hybrid Harmony Search and TRANSYT hill climbing algorithm for signalized stochastic equilibrium transportation networks. Transport. Res. C Emerg. Technol. 25, 152–167. https://doi.org/10.1016/j.trc.2012.05.007.

Cobos, C., Paz, A., Luna, J., Erazo, C., Mendoza, M., 2020. A multi-objective approach for the calibration of microscopic traffic flow simulation models. IEEE Access 8, 103124–103140. https://doi.org/10.1109/ACCESS.2020.2999081.

Creel, M., Loomis, J., 1997. Semi-nonparametric distribution-free dichotomous choice contingent valuation. J. Environ. Econ. Manag. 32 (3), 341–358. https://doi.org/10.1006/jeem.1997.0972.

Dia, H., Panwai, S., 2010. Evaluation of discrete choice and neural network approaches for modelling driver compliance with traffic information. [Article]. Transportmetrica 6 (4), 249–270. https://doi.org/10.1080/18128600903200596.

Diao, R., Shen, Q., 2012. Feature selection with harmony search. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 42 (6), 1509–1523. https://doi.org/10.1109/tsmcb.2012.2193613.

Emaasit, D., Paz, A., 2018. Simultaneous estimation of flexible models and associated hyperparameters: an application to travel activity-duration modeling. Transport. Res. Rec. 2672 (45), 147–159. https://doi.org/10.1177/0361198118788427.

Espino, R., Román, C., De Ortúzar, J.D., 2006. Analysing demand for suburban trips: a mixed RP/SP model with latent variables and interaction effects. Transportation 33 (3), 241–261. Retrieved from. https://www2.ulpgc.es/hege/almacen/download/46/46967/analysing_demand_for_suburban_trips_a_mixed_rpsp_model.pdf.

Fan, J., Li, R., 2006. Statistical challenges with high dimensionality: feature selection in knowledge discovery. In: International Congress of Mathematicians.

Fiebig, D.G., Keane, M.P., Louviere, J., Wasi, N., 2010. The generalized multinomial logit model: accounting for scale and coefficient heterogeneity. Market. Sci. 29 (3), 393–421. https://doi.org/10.1287/mksc.1090.0508.

Fosgerau, M., Bierlaire, M., 2007. A practical test for the choice of mixing distribution in discrete choice models. Transp. Res. Part B Methodol. 41 (7), 784–794. https://doi.org/10.1016/j.trb.2007.01.002.

Fosgerau, M., Hess, S., 2009. A comparison of methods for representing random taste heterogeneity in discrete choice models. European Transport - Trasporti Europei 42, 1–25. Retrieved from. https://eprints.whiterose.ac.uk/43631/.

Fountas, G., Sarwar, M.T., Anastasopoulos, P.C., Blatt, A., Majka, K., 2018. Analysis of stationary and dynamic factors affecting highway accident occurrence: a dynamic correlated grouped random parameters binary logit approach. Accid. Anal. Prev. 113, 330–340. https://doi.org/10.1016/j.aap.2017.05.018.

Gaudry, M., 1981. The inverse power transformation logit and dogit mode choice models. Transp. Res. Part B Methodol. 15 (2), 97–103. https://doi.org/10.1016/0191-2615(81)90036-9.

Geedipally, S.R., Lord, D., Dhavala, S.S., 2014. A caution about using deviance information criterion while modeling traffic crashes. Saf. Sci. 62, 495–498. https://doi.org/10.1016/j.ssci.2013.10.007.

Geem, Z.W., Kim, J.H., Loganathan, G.V., 2001. A new heuristic optimization algorithm: harmony search. Simulation 76 (2), 60–68. https://doi.org/10.1177/003754970107600201.

Goett, A.A., 1998. Estimating Customer Preferences for New Pricing Products Final Report. Retrieved from. http://inis.iaea.org/search/search.aspx?orig_q=RN:30046586.

Greene, W.H., Hensher, D.A., 2003. A latent class model for discrete choice analysis: contrasts with mixed logit. Transp. Res. Part B Methodol. 37 (8), 681–698. https://doi.org/10.1016/S0191-2615(02)00046-2.

Gundlach, A., Ehrlinspiel, M., Kirsch, S., Koschker, A., Sagebiel, J., 2018. Investigating people's preferences for car-free city centers: a discrete choice experiment. Transport. Res. Transport Environ. 63, 677–688. https://doi.org/10.1016/j.trd.2018.07.004.

Guo, Z., Shi, L., Chen, L., Liang, Y., 2017. A harmony search-based memetic optimization model for integrated production and transportation scheduling in MTO manufacturing. Omega 66, 327–343. https://doi.org/10.1016/j.omega.2015.10.012.

Han, Y., Zegras, C., Pereira, F.C., Ben-Akiva, M., 2020. A Neural-Embedded Choice Model: TasteNet-MNL Modeling Taste Heterogeneity with Flexibility and Interpretability. *arXiv preprint* Retrieved from. https://arxiv.org/abs/2002.00922.

Hensher, D.A., Greene, W.H., 2003. The Mixed Logit model: the state of practice. Transportation 30 (2), 133–176. https://doi.org/10.1023/A:1022558715350.

Hess, S., Rose, J.M., 2012. Can scale and coefficient heterogeneity be separated in random coefficients models? Transportation 39 (6), 1225–1239. https://doi.org/10.1007/s11116-012-9394-9.

Hess, S., Train, K., 2017. Correlation and scale in mixed logit models. Journal of Choice Modelling 23, 1–8. https://doi.org/10.1016/j.jocm.2017.03.001.

Kattan, A., Abdullah, R., Salam, R.A., 2010. Harmony search based supervised training of artificial neural networks. In: 2010 International Conference on Intelligent Systems, Modelling and Simulation, pp. 105–110, 27-29 Jan. 2010.

Keane, M., Wasi, N., 2013. Comparing alternative models of heterogeneity in consumer choice behavior. [Article]. J. Appl. Econom. 28 (6), 1018–1045. https://doi.org/10.1002/jae.2304.

Khadka, M., Paz, A., 2017. Comprehensive clusterwise linear regression for pavement management systems. J. Transport. Eng., Part B: Pavements 143 (4), 04017014. https://doi.org/10.1061/JPEODX.0000009.

Khadka, M., Paz, A., Arteaga, C., Hale, D.K., 2018. Simultaneous generation of optimum pavement clusters and associated performance models. Math. Probl Eng. 2018, 2159865 https://doi.org/10.1155/2018/2159865.

Kim, J., Rasouli, S., Timmermans, H., 2014. Hybrid choice models: principles and recent progress incorporating social influence and nonlinear utility functions. Procedia Environmental Sciences 22, 20–34. https://doi.org/10.1016/j.proenv.2014.11.003.

Kim, J., Rasouli, S., Timmermans, H., 2016. A hybrid choice model with a nonlinear utility function and bounded distribution for latent variables: application to purchase intention decisions of electric cars. Transportmetrica: Transport. Sci. 12 (10), 909–932. https://doi.org/10.1080/23249935.2016.1193567.

Kitazawa, Y., 2012. Hyperbolic transformation and average elasticity in the framework of the fixed effects logit model. Theor. Econ. Lett. 2 (No. 2), 192–199. https://doi.org/10.4236/tel.2012.22034.

Koppelman, F.S., 1981. Non-linear utility functions in models of travel choice behavior. Transportation 10 (2), 127–146. https://doi.org/10.1007/BF00165262.

Liu, Y.-H., Mahmassani, H.S., 2000. Global maximum likelihood estimation procedure for multinomial probit (MNP) model parameters. Transp. Res. Part B Methodol. 34 (5), 419–449. https://doi.org/10.1016/S0191-2615(99)00033-8.

Ma, Y., Zhang, Z., Ihler, A., 2020. A deep choice model for hiring outcome prediction in online labor markets. [Article]. Int. J. Comput. Commun. Control 15 (2), 15. https://doi.org/10.15837/ijccc.2020.2.3760.

Mandel, B., Gaudry, M., Rothengatter, W., 1997. A disaggregate Box-Cox Logit mode choice model of intercity passenger travel in Germany and its implications for high-speed rail demand forecasts. Ann. Reg. Sci. 31 (2), 99–120. https://doi.org/10.1007/s001680050041.

Mannering, F.L., Shankar, V., Bhat, C.R., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. Analytic Methods in Accident Research 11, 1–16. https://doi.org/10.1016/j.amar.2016.04.001.

Martín-Baos, J.Á., García-Ródenas, R., Rodriguez-Benitez, L., 2021. Revisiting kernel logistic regression under the random utility models perspective. An interpretable machine-learning approach. Transportation Letters 1–12. https://doi.org/10.1080/19427867.2020.1861504.

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

Ordoñez, C., Ruano, E., Cobos, C., Ordoñez, H., Ordoñez, A., 2018. Comparative analysis of MOGBHS with other state-of-the-art algorithms for multi-objective optimization problems. In: Paper Presented at Mexican International Conference on Artificial Intelligence, MICAI 2017. Advances in Soft Computingg, Cham.

Orro, A., Novales, M., Benitez, F., 2005. Nonlinearity and Taste Heterogeneity Influence on Discrete Choice Model Forecasts (Paper presented at European transport conference, Stasbourg, France).

Ortelli, N., Hillel, T., Pereira, F.C., 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.

Ortelli, N., Hillel, T., Pereira, F.C., Lapparent, M.d., Bierlaire, M., 2020. Assisted Specification of Discrete Choice Models. Retrieved from. https://transp-or.epfl.ch/documents/technicalReports/OrtHilPerLapBie2020.pdf.

Parady, G., Ory, D., Walker, J., 2021. The overreliance on statistical goodness-of-fit and under-reliance on model validation in discrete choice models: a review of validation practices in the transportation academic literature. Journal of Choice Modelling 38, 100257. https://doi.org/10.1016/j.jocm.2020.100257.

Paz, A., Arteaga, C., Cobos, C., 2019. Specification of mixed logit models assisted by an optimization framework. Journal of choice modelling 30, 50–60. https://doi.org/10.1016/j.jocm.2019.01.001.

Rahnamayan, S., Tizhoosh, H.R., Salama, M.M.A., 2008. Opposition-based differential evolution. IEEE Trans. Evol. Comput. 12 (1), 64–79. https://doi.org/10.1109/TEVC.2007.894200.

Ramsey, S.M., Bergtold, J.S., 2020. Examining inferences from neural network estimators of binary choice processes: marginal effects, and willingness-to-pay. [Article; early access]. Comput. Econ. 29 https://doi.org/10.1007/s10614-020-09998-w.

Revelt, D., Train, K., 2000. Customer-Specific Taste Parameters and Mixed Logit: Households' Choice of Electricity Supplier. UC Berkeley. Retrieved from. https://escholarship.org/uc/item/1900p96t.

Rodrigues, F., Ortelli, N., Bierlaire, M., Pereira, F., 2019. *arXiv preprint arXiv*:.03855. Bayesian Automatic Relevance Determination for Utility Function Specification in Discrete Choice Models. Retrieved from. https://arxiv.org/pdf/1906.03855.pdf.

Román, C., Arencibia, A.I., Feo-Valero, M., 2017. A latent class model with attribute cut-offs to analyze modal choice for freight transport. Transport. Res. Pol. Pract. 102, 212–227. https://doi.org/10.1016/j.tra.2016.10.020.

Ruano-Daza, E., Cobos, C., Torres-Jimenez, J., Mendoza, M., Paz, A., 2018. A multiobjective bilevel approach based on global-best harmony search for defining optimal routes and frequencies for bus rapid transit systems. Appl. Soft Comput. 67, 567–583. https://doi.org/10.1016/j.asoc.2018.03.026.

Sagebiel, J., Kirsch, S., Gundlach, A., Koschker, A., Ehrlinspiel, M., 2018. Data for: Investigating People's Preferences for Car-free City Centers: A Discrete Choice Experiment. Retrieved from. https://data.mendeley.com/datasets/p8t7pphvj3.

Schwarz, G., 1978. Estimating the dimension of a model. Ann. Stat. 6 (2), 461–464. Retrieved from. https://projecteuclid.org/download/pdf_1/euclid.aos/1176344136.

Sifringer, B., Lurkin, V., Alahi, A., 2018. Enhancing Discrete Choice Models with Neural Networks. In: 18th Swiss Transport Research Conference Monte Verità / Ascona. May 16 – 18.

Sifringer, B., Lurkin, V., Alahi, A., 2020. Enhancing discrete choice models with representation learning. Transp. Res. Part B Methodol. 140, 236–261. https://doi.org/10.1016/j.trb.2020.08.006.

Sillano, M., Ortúzar, J., 2005. Willingness-to-Pay estimation with mixed logit models: some new evidence. Environ. Plann.: Econ. Space 37 (3), 525–550. https://doi.org/10.1068/a36137.

Train, K., 2003. Discrete Choice Methods with Simulation. Cambridge University Press.

Train, K., 2007. A Recursive Estimator for Random Coefficient Models. University of California, Berkeley.

Train, K., 2016. Mixed logit with a flexible mixing distribution. Journal of Choice Modelling 19, 40–53. https://doi.org/10.1016/j.jocm.2016.07.004.

Train, K.E., 2008. EM Algorithms for nonparametric estimation of mixing distributions. Journal of Choice Modelling 1 (1), 40–69. https://doi.org/10.1016/S1755-5345(13)70022-8.

van Cranenburgh, S., Wang, S., Vij, A., Pereira, F., Walker, J., 2021. Choice modelling in the age of machine learning - discussion paper. Journal of Choice Modelling 100340. https://doi.org/10.1016/j.jocm.2021.100340.

Veeramisti, N., Paz, A., Khadka, M., Arteaga, C., 2021. A clusterwise regression approach for the estimation of crash frequencies. J. Transport. Saf. Secur. 13 (3), 247–277. https://doi.org/10.1080/19439962.2019.1611681.

Vij, A., Krueger, R., 2017. Random taste heterogeneity in discrete choice models: flexible nonparametric finite mixture distributions. Transp. Res. Part B Methodol. 106, 76–101. https://doi.org/10.1016/j.trb.2017.10.013 [Article].

Vij, A., Ryan, S., Sampson, S., Harris, S., 2018. Consumer preferences for mobility-as-a-service (MaaS) in Australia1.2018. In: 40th Australasian Transport Research Forum (ATRF).

Vinterbo, S., Ohno-Machado, L., 1999. A genetic algorithm to select variables in logistic regression: example in the domain of myocardial infarction. Proc AMIA Symp 984–988. Retrieved from. http://www.ncbi.nlm.nih.gov/pubmed/10566508.

Vrieze, S.I., 2012. Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). Psychol. Methods 17 (2), 228–243. https://doi.org/10.1037/a0027127.

Walker, J., Ben-Akiva, M., 2002. Generalized random utility model. Math. Soc. Sci. 43 (3), 303–343. https://doi.org/10.1016/S0165-4896(02)00023-9.

Wang, S., Wang, Q., Zhao, J., 2020a. Deep neural networks for choice analysis: extracting complete economic information for interpretation. Transport. Res. C Emerg. Technol. 118, 102701 https://doi.org/10.1016/j.trc.2020.102701.

Wang, S.H., Wang, Q.Y., Zhao, J.H., 2020b. Deep Neural Networks for Choice Analysis: Extracting Complete Economic Information for Interpretation, 118. Transportation Research Part C-Emerging Technologies, p. 22. https://doi.org/10.1016/j.trc.2020.102701 [Article].

Wang, X., Kim, S.H., 2019. Prediction and factor identification for crash severity: comparison of discrete choice and tree-based models. Transport. Res. Rec. 2673 (9), 640–653. https://doi.org/10.1177/0361198119844456.

Wu, H., Fai Cheung, S., On Leung, S., 2020. Simple use of BIC to assess model selection uncertainty: an illustration using mediation and moderation models. Multivariate Behav. Res. 55 (1), 1–16. https://doi.org/10.1080/00273171.2019.1574546.

Xiang, W., An, M., Li, Y., He, R., Zhang, J., 2014. An improved global-best harmony search algorithm for faster optimization. Expert Syst. Appl. 41 (13), 5788–5803. https://doi.org/10.1016/j.eswa.2014.03.016.

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.