Journal of Choice Modelling 47 (2023) 100419

Contents lists available at ScienceDirect

Journal of Choice Modelling

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

A Bayesian hierarchical approach to the joint modelling of Revealed and stated choices

Zili Li [a,b], Simon P. Washington [a,b], Zuduo Zheng [a,*], Carlo G. Prato [a,c]

[a] School of Civil Engineering, The University of Queensland, Brisbane, 4072, Australia
[b] Advanced Mobility Analytics Group Pty Ltd, Australia
[c] The School of Civil Engineering, the University of Leeds, UK

ARTICLE INFO

Keywords:
Bayesian hierarchical model
Discrete choice
Revealed preferences
Stated preferences
Travel mode choices

ABSTRACT

Revealed and stated choice data are fundamental inputs to understanding individuals' preferences. Owning to the distinctive characteristics and complementary nature of these two types of data, making joint inference based on their combined information content represents an attractive approach to preference studies. However, complications may arise from the different decision protocols under the two distinct choice contexts. In this study, a Bayesian hierarchical model is proposed to make joint inference from combined RP and SP data, with special attention paid to capturing the behavioural differences between the two choice contexts. In addition to the well-recognised issues of decision inertia and scale differences, the proposed model also takes into account other behavioural characteristics such as a decision-maker ignoring situation constraints, non-attending attributes, and misinterpreting attributes. An empirical analysis of a combined RP and SP dataset of travel mode choices is used to demonstrate the advantageous features of the model. Upon examining the empirical evidence, two main advantages emerge: the model provides direct measures of the effect of behavioural issues arising from ignoring situation constraints and non-attending attributes, as well as evidence for the misinterpretation of attributes.

1. Introduction

The modelling and inference of individuals' preferences are fundamental in many areas of research. In transport, the analysis of travellers' mode choices provides valuable information for designing policies, evaluating projects and managing infrastructure (e.g., Bhat 1997; Miller et al. 2005; Vij et al. 2013; Ye and Titheridge 2017). In marketing, the investigation of consumers' preferences for brands or products is central to predicting purchase intent and uncovering market dynamics (e.g., Guadagni and Little 1983; Andrews and Srinivasan 1995; Cobb-Walgren et al. 1995; Erdem and Keane 1996; Shin et al. 2012). In labour economics, the study of the factors related to work participation is fundamental for understanding workforce diversity (e.g., Baanders 2002; Broadway et al., 2017; Garcia et al., 2018). In health and environmental economics, the understanding of the public's attitudes and preferences is essential for evaluating health or environmental programs (e.g., Adamowicz et al., 1997; Whitehead et al., 2008; Mentzakis et al. 2011; Andersson et al. 2016).

In the aforementioned contexts, the collected data normally consist of observations of individual choices. Depending on how choice outcomes are obtained, data are typically categorised into two types: Revealed Preference (RP) data contain the observed choices of

* Corresponding author.
E-mail address: zuduo.zheng@uq.edu.au (Z. Zheng).

https://doi.org/10.1016/j.jocm.2023.100419
Received 19 December 2021; Received in revised form 19 January 2023; Accepted 12 May 2023
Available online 13 May 2023
1755-5345/

individuals making actual choices, whereas Stated Preference (SP) data contain the choices made by individuals under hypothetical and controlled situations. Numerous researchers claim that RP data are more reliable than SP data because they represent actual choice situations, whilst the validity of the inferred preferences based on SP data alone may be questionable because of their hypothetical nature (e.g., M. Ben-Akiva et al., 1994; Louviere et al., 2000; Hensher et al. 2005). It is worth noting that RP data can also be imperfect, for example decision-makers may not know the exact attribute values of non-chosen alternatives (e.g. the travel time of the non-chosen bus and the towing capacity of the non-chosen SUV). Despite the known limitations of SP data, they have played and will continue to play an important role in preference studies and are accepted as a valid method for understanding preference behaviour. Their popularity is related to having a complementary nature to RP data, an ability to explicitly vary attribute values across alternatives, an advantage derived from designed experiments typical of SP studies, and being an ideal method for obtaining information on preferences for currently unavailable market offerings.

Due to the complementary nature of RP and SP data, making joint inferences from combined RP and SP data has emerged as an attractive approach to preference studies (M. Ben-Akiva et al., 1994; Brownstone et al. 2000; Bhat and Castelar 2002; Hensher et al. 2008; Cherchi and Juan de Dios Ortúzar, 2011). Combining RP and SP data allows inferences to be made based on all available information and leads to increased sample size and outcome robustness. However, additional challenges arise from making joint inferences, with two widely recognised issues being decision inertia and scale difference (Moshe Ben-Akiva and Morikawa, 1990; Bradley and Daly 1997; Brownstone et al. 2000; Morikawa et al. 2002; Cherchi and Juan de Dios Ortúzar, 2006; Börjesson 2008). Decision inertia — or state dependence —refers to the tendency of individuals to repeat the same choice: for example, SP choices may be influenced by familiarity with previously chosen alternatives or repetition of recent RP choices. Scale differences refer to the magnitude of utility coefficients being different: for example, SP estimated parameters may differ from RP estimated parameters not only because of an actual difference, but because of differences in the number of omitted factors affecting the choice across the RP and SP contexts — thus affecting the scale of the parameter estimates.

The decision inertia of decision-makers is usually captured within the utility function by incorporating an indicator variable that, for example, takes value one if the current choice in the SP data was also chosen by the same individual in the RP data — and essentially enables the testing of significance of the influence of one choice on the other. In contrast, scale difference is usually captured by normalising to the scale of the RP data and estimating de facto the ratio of the scale parameters that captures whether the SP data have less or more variance. While the consideration of decision inertia and scale difference represents an important aspect of preference studies from combining RP and SP data, additional features should also be considered when making joint inference.

Conceptually, the aforementioned issues arise from one fundamental source — the differences in choice behaviour under contrasting situations. For example, decision inertia arises when an individual's decision protocol under hypothetical situations minimises the cost of evaluating unfamiliar information — instead relying more heavily on past choice behaviour. In another example, scale differences arise when different sets of attributes are employed for uncovering preferences. Both of these examples highlight the underpinning nature of preference studies, namely that the heterogeneous behaviour of individuals requires careful capturing and modelling of complex decision making.

However, these two issues do not cover the gamut of issues from combining SP and RP data, and the present study contributes to the literature by delving into the details of how behavioural differences in the two contrasting choice contexts can be better captured in a model. Specifically, the present study combines serval well-established modelling techniques and proposes a modelling framework that accounts not only for the vastly investigated issues of decision inertia and scale differences, but also for the largely overlooked issues of decision-makers ignoring situation constraints, non-attending attributes and misinterpreting attributes. These issues have been long suspected to emerge from the two choice contexts on a conceptual level (Morikawa et al. 2002; Ben-Akiva et al., 2019), but have not been tackled in an integrated statistical model.

The remainder of the paper is organised as follows. Section 2 describes the behavioural constructs and underpinnings that motivate the model. Section 3 introduces the utility functions and the Bayesian estimation procedure of the model. Section 4 illustrates the application of the proposed model using a joint RP and SP travel mode choice dataset. Section 5 draws conclusions from the study and proposes further research avenues.

2. The conceptual framework

Given the differences between the RP and the SP choice contexts, any model seeking to make joint inferences from a combined dataset should account for the potential deviations of individuals' choice behaviour from their underlying preference. With this premise, a natural starting point for the present study is to look at the behavioural perspective by identifying major differences and their potential impact on inferred preferences, and the modelling perspective by designing a strategy that addresses unexplored issues.

2.1. Behavioural perspective

From the behavioural perspective, five potential discrepancies have been identified to exist between the decision protocols in SP and RP contexts (Morikawa et al. 2002).

1) Respondents tend to make the same choice.
2) Questionnaires are used as opinion statements.
3) Situation constraints are ignored.
4) Respondents only consider the most important attributes.

5) Attributes are misinterpreted.

Given these discrepancies, the preferences inferred from SP contexts could differ from those extracted from RP contexts, and the estimated parameters from SP choices (and hence from combined SP and RP choices) could be biased. Consequently, an important feature of a joint modelling framework is the ability to account for the potential biases arising from these discrepancies. Arguably, a joint modelling framework would be desirable that can disentangle the possible biases originated from the different discrepancies.

Let us first examine the five discrepancies. Decision inertia corresponds clearly to the first one, whereas scale difference corresponds reasonably to a blending of the remaining four. The second discrepancy describes bias that might arise from respondents overstating their preference for one alternative over the others in order to promote its acceptance, support or realisation. The third discrepancy relates to bias from a hypothetical choice construct; as examples, a decision-maker does not experience congestion in a hypothetical route choice nor does a decision-maker bear the financial burden in a hypothetical vehicle purchase. The last two discrepancies describe possible biases that might arise from respondents avoiding the high cost of acquiring information about alternatives in RP and SP contexts, and difference in included and omitted attributes across both contexts. Accordingly, the attributes considered in the SP and RP contexts may be different across decision-makers because of behavioural and specification reasons: (i) influential attributes may differ across choice contexts from a behavioural perspective; (ii) an SP choice experiment may include irrelevant but covarying attributes, or may omit important attributes. The potential impact of non-attended attributes are typically absorbed by the model error term, and hence a variance (scale) difference for the error terms would emerge in addition to differences in the magnitudes of estimated model parameters.

A desirable joint modelling framework should address the use of questionnaires as opinion statements to correct for biases towards an alternative, should consider situation constraints that affect choices in the RP situations (but are ignored in the SP context), and should correct for the omission and/or misinterpretation of attributes when the model specification across the SP and RP attribute sets are different. In a nutshell, the list of discrepancies provides a roadmap for the modelling strategy presented in the following section — and is the motivation for exploring the model specification and empirical testing described in this paper.

2.2. Modelling perspective

From the modelling perspective, the list of behavioural discrepancies translates to model features that aim to capture biases and scale differences. Given this behavioural roadmap, we describe a modelling approach that requires four features.

1) An indicator variable is used to capture decision inertia.
2) A scale parameter is used to capture scale difference.
3) Alternative-specific intercepts are used to allow for adjustments of bias that arise from decision-makers using questionnaires as opinion statements and from ignoring situation constraints.
4) Parameter expansion for all attributes as a product of a random indicator parameter and a magnitude parameter allows adjustments of bias from non-attendance and/or misinterpretation of attributes.

The first three model features are extracted from the existing literature. The first feature introduces an indicator that captures the decision inertia in the SP situations. The second feature establishes an indicator that accounts for differences in the variance of the error term and the unobserved utilities between the SP and RP data. And the third feature uses alternative-specific intercepts with the idea of correcting the bias for alternatives that are chosen to promote their acceptance and realisation (using questionnaires as opinion statements), as well as bias that might arise from ignoring the situation constraints not present in SP contexts. Specifically, the intercepts correct for the excess utility arising from these two biases.

The last model feature allows for the expansion of each parameter to capture additional choice complexity — akin to how a zero inflated model captures a dual state process in count data. To do this, an individual-specific indicator parameter captures the existence of influence of the attributes, while a magnitude parameter captures the level of non-zero attribute effects. The indicator parameter is estimated to correct for the bias from non-attendance of attributes, by estimating not only which attributes have been considered in the choice contexts, but also to which extent. Specifically, a prior distribution is assigned to all individual-specific indicator parameters in a Bayesian hierarchical framework to capture the relative importance of each attribute in terms of the probability of being considered in the choice context. The magnitude parameter captures the non-zero effect of each attribute and corrects for the bias from misinterpreting the attribute by separating non-attendance from the non-zero effect of the parameter. From a methodological point of view, the way of using indicator variables is similar to that in Bayesian variables selection (O'Hara and Sillanpää 2009). However, it will be demonstrated that important and novel interpretations regarding the behaviour of respondents can be derived when it is used in the decision-making contexts.

To summarise, the model contains four features that correct for the biases from the five behavioural discrepancies: the first feature corrects for choice inertia, the second feature captures scale differences, and the third feature captures situation constraints, whilst the fourth feature captures attribute non-attendance and misinterpretation.

3. Model specification

The foundational model specification is the mixed logit commonly used for discrete choice modelling (McFadden 1974). In both the frequentist and Bayesian approaches, the probability P_{nit} of individual n choosing alternative i on choice occasion t is expressed as a function of unknown parameters $\boldsymbol{\beta}$ and explanatory variables $\boldsymbol{X}$. Given the dataset $\mathscr{D} = \{\boldsymbol{Y}, \boldsymbol{X}\}$ that includes choice outcomes $\boldsymbol{Y}$, the probability $P(\boldsymbol{Y}|\boldsymbol{\beta}, \boldsymbol{X})$ of the observed sequence of all individuals' choice outcomes in all choice occasions given $\mathscr{D}$ and $\boldsymbol{\beta}$ is expressed as follows:

$$P(\boldsymbol{Y}|\boldsymbol{\beta}, \boldsymbol{X}) = \prod_{n=1}^{N} \prod_{t=1}^{T} \prod_{i=1}^{I} [P_{nit}(\boldsymbol{Y}|\boldsymbol{\beta}, \boldsymbol{X})]^{y_{nit}} \tag{1}$$

where y_{nit} is a choice indicator variable (equal to one when individual n chooses alternative i in choice occasion t, and zero otherwise), N is the number of individuals, T is the number of choice occasions, and I is the number of alternatives, respectively.

Furthermore, the probability $P_{nit}(\boldsymbol{Y}|\boldsymbol{\beta}, \boldsymbol{X})$ of individual n choosing alternative i on choice occasion t is expressed as:

$$P_{nit}(\boldsymbol{Y}|\boldsymbol{\beta}, \boldsymbol{X}) = \frac{e^{V_{nit}(\boldsymbol{\beta}, \boldsymbol{X})}}{\sum_{i=1}^{I} e^{V_{nit}(\boldsymbol{\beta}, \boldsymbol{X})}} \tag{2}$$

where $V_{nit}(\boldsymbol{\beta}, \boldsymbol{X})$ is the deterministic part of the utility function. The functional form of $V_{nit}(\boldsymbol{\beta}, \boldsymbol{X})$ may be assumed to be linear:

$$V_{nit}(\boldsymbol{\beta}, \boldsymbol{X}) = \boldsymbol{\beta}_{\boldsymbol{ni}}^{'} \boldsymbol{X}_{\boldsymbol{nit}} \tag{3}$$

where the elements of column vector $\boldsymbol{\beta}_{\boldsymbol{ni}}$ are parameters that are specific for individual n and alternative i, and the elements of column vector $\boldsymbol{X}_{\boldsymbol{nit}}$ are the corresponding explanatory variables that are specific for individual n and alternative i on occasion t. The size of both column vectors equal to M+1, where M is the number of attributes.

3.1. The utility function

Eqs. (1)–(3) as a whole form the core of the likelihood function of the observed choices. In light of the conceptual background from the behavioural and modelling perspective, modifications to the utility function in eq. (3) are made to accommodate the aforementioned four behavioural features.

The first two features correct for decision inertia and scale differences (Brownstone et al. 2000; Bhat and Castelar 2002), and the utility function $V_{nit}(\boldsymbol{\beta}, \boldsymbol{X})$ is modified from eq. (3) by considering scale parameters λ_{nt} and inertia parameters θ_{ni}:

$$V_{nit}(\boldsymbol{\beta}, \boldsymbol{X}) = \lambda_{nt} v_{nit}(\boldsymbol{\beta}, \boldsymbol{X}) \tag{4}$$

$$\lambda_{nt} = D_{nt}^{RP} + \lambda\left(1 - D_{nt}^{RP}\right) \tag{5}$$

$$v_{nit}(\boldsymbol{\beta}, \boldsymbol{X}) = \boldsymbol{\beta}_{\boldsymbol{ni}}^{'} \boldsymbol{X}_{\boldsymbol{nit}} + \theta_{ni}\left(1 - D_{nt}^{RP}\right)\mathbf{1}_{\left\{\sum_{t=1}^{T} D_{nt}^{RP} y_{nit} > 0\right\}} \tag{6}$$

In eq. (5), D_{nt}^{RP} is an indicator variable (equal to one if the observation of individual n on occasion t is made in an RP context, and zero otherwise), and λ is a parameter that allows for the scale of the utility to deviate from one for the RP data to λ for the SP data. In eq. (6), the utility $v_{nit}(\boldsymbol{\beta}, \boldsymbol{X})$ of alternative i for individual n on occasion t in SP contexts is adjusted by an amount equal to θ_{ni} if individual n has chosen alternative i at least once in the RP situation (as expressed by the term $\mathbf{1}_{\{\sum_{t=1}^{T} D_{nt}^{RP} y_{nit} > 0\}}$). An alternative method for dealing with decision inertia could be the use of autocorrelated error terms as in Allenby and Lenk (1994).

The third model specification feature corrects for the bias from using questionnaires as opinion statements and ignoring situation constraints. Both behavioural discrepancies are at the root of possible bias that might shift upwards the latent utility of the affected alternatives with respect to the utilities of the unaffected alternatives, regardless of the values of the attributes considered in the utility function. These shifts in the latent utilities reflect changes in the values of the intercepts when making choices across the RP to the SP contexts, and importantly, these changes can affect not only the relative scale of the latent utilities, but also their ordering. The proposed model accounts for these changes with intercepts that (i) are allowed to vary based on the observation being in the RP or the SP data, and (ii) are alternative-specific. The utility function $v_{nit}(\boldsymbol{\beta}, \boldsymbol{X})$ is then modified as follows:

$$\begin{aligned} v_{nit}(\boldsymbol{\beta}, \boldsymbol{X}) &= \beta_{0ni}^{RP} D_{nt}^{RP} + \beta_{0ni}^{SP}\left(1 - D_{nt}^{RP}\right) + \boldsymbol{\beta}_{\boldsymbol{1ni}}^{'} \boldsymbol{X}_{\boldsymbol{nit}} \\ &+ \theta_{ni}\left(1 - D_{nt}^{RP}\right)\mathbf{1}_{\left\{\sum_{t=1}^{T} D_{nt}^{RP} y_{nit} > 0\right\}} \end{aligned} \tag{7}$$

where the elements of the vector $\boldsymbol{\beta}_{\boldsymbol{ni}}$ are separated into two components: the intercepts and the parameters associated with the explanatory variables (the attributes). The intercepts are further decomposed into the individual and alternative-specific intercepts β_{0ni}^{RP} and β_{0ni}^{SP} (for the RP and SP data, respectively). The remaining parameters associated with the attributes are elements of the $M\times$ 1 vector $\boldsymbol{\beta}_{\boldsymbol{1ni}}$.

The fourth model specification feature corrects for the bias arising from non-attendance or misinterpretation of attributes. An attribute might be considered only in the RP or SP contexts, or both, although its impact might be different across the RP and SP contexts. It should be noted that attributes might capture situation constraints as well: imagine for example *income* being captured in both RP and SP data, but being more influential in RP choices than in hypothetical SP choices. The proposed model accounts for the possible non-attendance and misinterpretation of attributes by redefining each parameter for each attribute as the product of an indicator parameter (capturing the attendance of the attribute in the individual's choice process) and a magnitude parameter (isolating the effect of the corresponding attribute):

$$v_{nit}(\boldsymbol{\beta},\boldsymbol{X}) = \alpha_{0i}^{RP} I_{0n}^{RP} D_{nt}^{RP} + \alpha_{0i}^{SP} I_{0n}^{SP}\left(1 - D_{nt}^{RP}\right) + \boldsymbol{\alpha}_{1i}^{'} \boldsymbol{I}_n \boldsymbol{X}_{nit}$$
$$+\alpha_{\theta i} I_{\theta n}\left(1 - D_{nt}^{RP}\right) \mathbf{1}_{\left\{\sum_{t=1}^{T} D_{nt}^{RP} y_{nit} > 0\right\}} \tag{8}$$

where the intercepts β_{0ni}^{RP} and β_{0ni}^{SP} are expressed as the product of the respective indicator parameters I_{0n}^{RP} and I_{0n}^{SP} (taking value zero or one) and the respective magnitude parameters α_{0i}^{RP} and α_{0i}^{SP}, the parameters associated with the attributes formulated as the product of a $M \times M$ diagonal matrix $\boldsymbol{I}_n$ of indicator parameters and the respective $M \times 1$ vector $\boldsymbol{\alpha}_{1i}$ of magnitude parameters for all the attributes, and the inertia is captured with the product of the inertia indicators $I_{\theta n}$ and the inertia magnitudes $\alpha_{\theta i}$.[1] It should be noted that, for notational simplicity, $\boldsymbol{\alpha}_{1i}$ and $\boldsymbol{I}_n$ denote vectors and matrices containing both the RP and the SP parameters: the number of the corresponding explanatory variables and the observations for each variable are hence doubled approximately, and the variables corresponding to an RP-specific parameter have value zero for SP observations, and vice versa.

Notably, combining the utility function in eq. (8) with the probability formulations in eqs. (1) and (2) provides the probability of the observed sequence of RP and SP choices while having all the four model features incorporated to capture the five behavioural discrepancies.

3.2. Prior specifications

The proposed model is estimated in a Bayesian framework that requires the specification of prior distributions. A hierarchical prior structure is also needed to shift the focus of the inference since preference studies usually focus on the population rather than the individual level.[2] Moreover, a hierarchical prior over individual specific parameters allows us to impose parameter restriction that is beneficial for both model estimation and inference.

The prior structure applies to the population distributions for the context-specific magnitude parameters (α_{0i}^{RP}, α_{0i}^{SP}, $\boldsymbol{\alpha}_{1i}$, and $\alpha_{\theta i}$) and individual-specific indicator parameters (I_{0n}^{RP}, I_{0n}^{SP}, $\boldsymbol{I}_N$, and $I_{\theta n}$) in the utility function $v_{nit}(\boldsymbol{\beta},\boldsymbol{X})$ presented in eq. (8). It should be noted that we refer to a mixed logit model, in that the population distribution is obtained from the individual-specific or context-specific parameters rather than the traditional parameterisation most commonly described in the literature.

For the context-specific magnitude parameters, the idea is to infer the effect of any attribute on individuals' decision process under the assumption that there is a certain level of attendance of the attributes captured by the indicator parameters. The magnitude parameters for each attribute are assumed to be different across the RP and SP data, although it is possible that they are invariant in cases when behavioural discrepancies do not manifest across the two contexts. The priors are assumed not only to be different for the RP and SP data, but also specific to each attribute and each alternative. Moreover, the following distributions are assumed for population parameters: normal distributions ($\mathscr{N}$) for real valued parameters, inverse gamma distributions ($\mathscr{I}\mathscr{G}$) with the scale parameter for positive real valued parameters, and beta distributions (*Beta*) for parameters taking values between zero and one. These distributions are the standard conjugate priors, and other alternative distributions can be found in the works of Gelman (2006), Gelman et al. (2013) and Huang and Wand (2013).

For the individual-specific indicator parameters, the aim is to infer whether an attribute is attended in an individual's decision process. The indicator parameters for each individual and each attribute are assumed to be the same across alternatives: two indicator parameters are generated for each individual from two independent population distributions, one for RP data and the other for the SP data. This specification accommodates potential differences in the sets of attributes that could be related to the individuals' choices under different situations, thus correcting for not only attribute non-attendance, but also situation constraints. The priors are assumed to be independent across the RP and SP data, and Bernoulli distributions ($\mathscr{B}$) are used as the population distributions of the two indicator parameters.

Given the model specification and desire for computational efficiency, the hierarchical priors in the proposed model take the following specification:

$$p_{md} \sim Beta(a_{md}^{p}, b_{md}^{p}) \quad m \in \{0,\ldots,M\}, d \in \{RP, SP\} \tag{9}$$

[1] In a related study (Gilbride and Allenby 2004), similar indicator parameters are used for the purposes of selecting alternatives instead of the attributes of alternatives. In our case, the indicator parameter associated with each attribute is used to identify the existence of the impact of that particular attribute on the probability of the chosen alternative, whereas in their case, the indicator parameter is included to identify the existence of the effect of each alternative's utility on the chosen alternative.

[2] Henceforth, population distribution will refer to the distribution of individual-specific or context specific parameters: "population" does not mean all the individuals in a sample or in any larger scope.

$$\mu_{0di} \sim \mathcal{N}(\mu^{\mu}_{0di}, \sigma^{\mu}_{0di}) \quad d \in \{RP, SP\}, i \in \{1, \ldots, I\} \tag{10}$$

$$\sigma_{0di} \sim \mathcal{IG}\left(a^{\sigma}_{0di}, b^{\sigma}_{0di}\right) \quad d \in \{RP, SP\}, i \in \{1, \ldots, I\} \tag{11}$$

$$\mu_{mi} \sim \mathcal{N}(\mu^{\mu}_{mi}, \sigma^{\mu}_{mi}) \quad m \in \{1, \ldots, M\}, i \in \{1, \ldots, I\} \tag{12}$$

$$\sigma_{mi} \sim \mathcal{IG}\left(a^{\sigma}_{mi}, b^{\sigma}_{mi}\right) \quad m \in \{1, \ldots, M\}, i \in \{1, \ldots, I\} \tag{13}$$

$$\lambda \middle| a^{\lambda}, b^{\lambda} \sim \mathcal{IG}\left(a^{\lambda}, b^{\lambda}\right) \tag{14}$$

$$I^{d}_{mn} \sim \mathcal{B}(p_{md}) \quad m \in \{0, \ldots, M\}, d \in \{RP, SP\}, n \in \{1, \ldots, N\} \tag{15}$$

$$\alpha^{d}_{0i} \sim \mathcal{N}(\mu_{0di}, \sigma_{0di}) \quad d \in \{RP, SP\}, i \in \{1, \ldots, I\} \tag{16}$$

$$\alpha^{d}_{mi} \sim \mathcal{N}(\mu_{mi}, \sigma_{mi}) \quad m \in \{1, \ldots, M\}, d \in \{RP, SP\}, i \in \{1, \ldots, I\} \tag{17}$$

where m is the index for the $M + 1$ attributes including the intercept ($m = 0$) and the inertia indicators, and d is the index for the RP and SP data. With the exception of the use of d to indicate RP or SP data, the superscripts indicate the hyper-prior parameters, namely the parameters of the population distributions, which are specified as weakly informative.

The first block of the prior specification (eqs. (9)–(14)) describes the population parameters to be drawn. These parameters include: the probability parameter p_{md} for each attribute m and data type d; the parameters for the population distributions of the two (RP and SP) intercepts of each alternative i, each with mean μ_{0di} and standard deviation σ_{0di}; the parameters for the population distributions of the magnitude parameters of all attributes or inertia indicator m of each alternative i, each with mean μ_{mi} and standard deviation σ_{mi}; and the scale parameter λ.

The second block of the prior specification (eqs. (15)–(17)) defines the individual-specific or context-speicfic parameters that are drawn on the basis of the population parameters from the first block. These parameters include: the indicator parameters I^{d}_{mn} for each attribute m and individual n of data type d, which are conditional on the probability parameters p_{md}; the magnitude parameters α^{d}_{0i} of the intercepts of each alternative i and data type d; the magnitude parameters α^{d}_{mi} of each attribute or inertia indicator m of each alternative i and data type d.

In this prior structure, it should be noted that one distinct population distribution is assumed for the magnitude parameters of each attribute m and each alternative i, and this distribution is the same for the RP and SP data with the exception of the intercepts. This allows for the parameters of the population distributions to be jointly estimated from the choices made in both the RP and SP contexts, and for the underlying population preference over the attributes to be inferred. Notably, the different distributions for the intercepts enable the model to account for the possible bias arising from behavioural discrepancies such as using questionnaires as opinion statements and ignoring situation constraints.

Finally, the population distributions for indicator parameters are different across the RP and SP contexts, but the same across alternatives. This allows for the parameters of the population distribution to be jointly estimated to infer the importance of attributes in individuals' decision protocols in the two different choice contexts, regardless of which specific alternative is affected by the attributes. Notably, these distributions enable the model to account for attendance levels of attributes, misinterpretation of attributes, and ignored situation constraints.

3.3. Posterior sampling

Considering the utility function $v_{nit}(\boldsymbol{\beta}, \boldsymbol{X})$ in eq. (8) and the likelihood function $P(\boldsymbol{Y}|\boldsymbol{\beta}, \boldsymbol{X})$ of the observed sequence of all individuals' choice outcomes in all choice occasions, the joint posterior for all the unknown parameters $\boldsymbol{\beta}$ given the data $\mathcal{D}$ is defined as:

$$P(\boldsymbol{\beta}|\boldsymbol{Y}, \boldsymbol{X}) \propto P(\boldsymbol{Y}|\boldsymbol{\beta}, \boldsymbol{X}) \times P_{\beta_n}\left(\boldsymbol{\beta_n}\middle|\boldsymbol{\beta_p}\right) \times P_{\beta_p}\left(\boldsymbol{\beta_p}\middle|\boldsymbol{\beta_0}\right) \tag{18}$$

where the population parameters $\boldsymbol{\beta_p}$ have joint probability $P_{\beta_p}(\boldsymbol{\beta_p}|\boldsymbol{\beta_0})$ conditional on the hyper-prior parameters $\boldsymbol{\beta_0}$, and the individual level parameters $\boldsymbol{\beta_n}$ have joint probability $P_{\beta_n}(\boldsymbol{\beta_n}|\boldsymbol{\beta_p})$ conditional on the population parameters $\boldsymbol{\beta_p}$. Accordingly, the parameters $\boldsymbol{\beta}$ include both $\boldsymbol{\beta_n}$ and $\boldsymbol{\beta_p}$:

$$\boldsymbol{\beta} = \{\boldsymbol{\beta_n}\} \cup \{\boldsymbol{\beta_p}\} \tag{19}$$

Consider the hyper-prior parameters $\boldsymbol{\beta_0}$:

$$\boldsymbol{\beta_0} = \{a^{p}_{md}, b^{p}_{md}; m \in \{0, \ldots, M\}, d \in \{RP, SP\}\} \cup \left\{\mu^{\mu}_{0di}, \sigma^{\mu}_{0di}, a^{\sigma}_{0di}, b^{\sigma}_{0di}; d \in \{RP, SP\}, i \in \{1, \ldots, I\}\right\} \cup \left\{\mu^{\mu}_{mi}, \sigma^{\mu}_{mi}, a^{\sigma}_{mi}, b^{\sigma}_{mi}; m \in \{1, \ldots, M\}, i \in \{1, \ldots, I\}\right\} \cup \left\{a^{\lambda}, b^{\lambda}\right\} \tag{20}$$

Then, the probability $P_{\beta_p}(\boldsymbol{\beta_p}|\boldsymbol{\beta_0})$ of the population parameters is calculated conditional on the hyper-prior parameters $\boldsymbol{\beta_0}$:

$$P_{\beta_p}(\boldsymbol{\beta_p}|\boldsymbol{\beta_0}) = \prod_{\substack{m\in\{0,\dots,M\}\\ d\in\{RP,SP\}}} Beta(p_{md}|a^p_{md}, b^p_{md}) \times \prod_{\substack{d\in\{RP,SP\}\\ i\in\{1,\dots,I\}}} \left(\mathscr{N}(\mu_{0di}|\mu^\mu_{0di}, \sigma^\mu_{0di}) \mathscr{IG}\left(\sigma_{0di}|a^\sigma_{0di}, b^\sigma_{0di}\right)\right) \times \prod_{\substack{m\in\{1,\dots,M\}\\ i\in\{1,\dots,I\}}} \left(\mathscr{N}(\mu_{mi}|\mu^\mu_{mi}, \sigma^\mu_{mi}) \mathscr{IG}\left(\sigma_{mi}|a^\sigma_{mi}, b^\sigma_{mi}\right)\right) \times \mathscr{IG}\left(\lambda|a^\lambda, b^\lambda\right) \tag{21}$$

where the population parameter $\boldsymbol{\beta_p}$ are:

$$\boldsymbol{\beta_p} = \{p_{md}; m\in\{0,\dots,M\}, d\in\{RP,SP\}\} \cup \{\mu_{0di}, \sigma_{0di}; d\in\{RP,SP\}, i\in\{1,\dots,I\}\} \cup \{\mu_{mi}, \sigma_{mi};\ m\in\{1,\dots,M\}, i\in\{1,\dots,I\}\} \cup \{\lambda\} \tag{22}$$

Then, the probability $P_{\beta_n}(\boldsymbol{\beta_n}|\boldsymbol{\beta_p})$ of the individual level parameters is calculated conditional on the population parameters $\boldsymbol{\beta_p}$:

$$P_{\beta_n}(\boldsymbol{\beta_n}|\boldsymbol{\beta_p}) = \prod_{\substack{m\in\{0,\dots,M\}\\ d\in\{RP,SP\}\\ n\in\{0,\dots,N\}}} \mathscr{B}\left(I^d_{mn}|p_{md}\right) \times \prod_{\substack{d\in\{RP,SP\}\\ i\in\{1,\dots,I\}}} \mathscr{N}\left(\alpha^d_{0i}|\mu_{0di}, \sigma_{0di}\right) \times \prod_{\substack{m\in\{1,\dots,M\}\\ d\in\{RP,SP\}\\ i\in\{1,\dots,I\}}} \mathscr{N}\left(\alpha^d_{mi}|\mu_{mi}, \sigma_{mi}\right) \tag{23}$$

where the individual level parameters $\boldsymbol{\beta_n}$ are:

$$\boldsymbol{\beta_n} = \left\{I^d_{mn}; m\in\{0,\dots,M\}, d\in\{RP,SP\}, n\in\{1,\dots,N\}\right\} \cup \left\{\alpha^d_{0i};\ d\in\{RP,SP\}, i\in\{1,\dots,I\}\right\} \cup \left\{\alpha^d_{mi}; m\in\{1,\dots,M\}, d\in\{RP,SP\}, i\in\{1,\dots,I\}\right\} \tag{24}$$

Lastly, conditional on the individual level parameters $\boldsymbol{\beta_n}$ and the population parameters $\boldsymbol{\beta_p}$, the probability of the sequence of the observed choice outcomes is given by substituting the values of the parameters in the utility function $v_{nit}(\boldsymbol{\beta}, \boldsymbol{X})$ in eq. (8) and then using eqs. (2) and (1).

From the computational perspective, the joint posterior distribution of all the unknown parameters $\boldsymbol{\beta}$ in eq. (19) can be approximated by using Markov Chain Monte Carlo sampling (Hastings 1970; Gelfand and Smith 1990). It might appear that the posterior sampling is computationally demanding since the posterior distribution is a product of functions for all individuals, all alternatives, all attributes, and both choice contexts. However, only a small portion of the above expressions change depending on the specific parameters being sampled in the sampling process. For example, the Metropolis algorithm (Gelman et al., 2013) can be used to sample $\boldsymbol{\beta_n}$ for one individual at a time, conditional on the parameters for the other individuals. Consequently, the ratio of the posterior probability of the proposed move with respect to the current position would depend only on the probability for that individual and the corresponding population distribution of the parameters being sampled. Moreover, Gibbs steps can be used in the sampling of all the population parameters ($\boldsymbol{\beta_p}$) on the basis of the closed form conditional posterior distributions (Gelman et al., 2013).

These considerations reveal that, in addition to the behavioural motivations at the root of the proposed model, computational considerations apply. Notably, the proposed model uses individual-specific indicator parameters, leading to the formulation of each parameter for each attribute as the product of a magnitude and an indicator parameter. Essentially, this approach aligns with the commonly used Bayesian variable selection method (Kuo and Mallick 1998; Smith and Kohn 2002; Sillanpää and Bhattacharjee 2005). However, this approach differs from the traditional variable selection in that the indicator parameters in the proposed model are individual-specific rather than common to all observations. This difference is appealing not only from the behavioural perspective, as intuitively there are anticipated differences across individuals in the decision protocols and the set of attributes considered, but also from the methodological perspective, as it is essential for the successful implementation of the proposed model.

In fact, when the indicator parameters are assumed to be the same across individuals as in the traditional Bayesian variable selection literature, it is found that the values of the indicators hardly change (O'Hara and Sillanpää 2009). The reason is that, once the indicator takes value zero, the proposed moves of the corresponding magnitude parameters would be evaluated entirely on the basis of the probability determined by other magnitude parameters with the corresponding indicators taking value one. Consequently, the Markov chains do not tend to stay within the high posterior probability region of those magnitude parameters with the corresponding indicators taking value zero. In turn, the relevant indicators would hardly flip back to one, resulting in a poor mixing of the chains. The proposed model largely alleviates this issue because of the use of population distributions that allow the proposed moves of the magnitude parameters for the individuals having the corresponding indicators equal to zero to be guided by other individuals with corresponding indicators equal to one. Moreover, using individual-specific indicators allows for a better mixing of the chains because the ratio between the posterior probability of indicator parameters at the proposed value P^p_n and at the current value P^c_n would be equal to P^p_n/P^c_n rather than being a product of the ratios for all individuals.

Moreover, the proposed model considers the difference in the amount of observations for the SP and the RP data. Typically, the SP part of a dataset contains a number of choice outcomes for each individual in several scenarios that are each represented by a different set of values for the SP-specific attributes, while the RP part of a dataset represents one scenario. This feature leads to a much higher number of observations in the SP part of the data and leads to parameter estimates for the attributes that are in both the RP and SP situations to be determined mainly on the SP choices. Although this might not be an issue depending on prior beliefs, the proposed model re-weighs the likelihood function so that for each individual the total number of the RP choices will have the same weight as the total number of the SP choices. Accordingly, the re-weighing leads to the joint inference on parameter estimates being determined equally across RP and SP observations. Although arbitrary, the re-weighing may be considered as a way of reflecting the prior belief of the modeller and may be seen as a natural fit into the Bayesian framework where the posterior distribution obtained from the RP data is used as a prior belief (distribution) for the model of the SP data. Then, the strength of this prior is represented by the weight that can be interpreted as the credibility of each RP observation in comparison to that of each SP response.

4. An application to travel mode choice

Given the behavioural motivation and the model specification, the proposed model was estimated on a joint RP and SP dataset of travel mode choices. In this application, an initial 10,000 iterations were done as the burn-in period. During the burn-in period, the proposal distribution of the sampler is updated periodically based on previous iterations. Another 5000 iterations after the burn-in period are used as the approximation to the posterior distribution for inference purposes. The convergence of the chain is judged by using trace plots of the sampled parameters and the values of log-likelihood.

From a behavioural standpoint, the term *attribute* is used for representing both a characteristic of an alternative and that of an individual in this section. However, the attendance or non-attendance of an attribute is used exclusively in the context of alternatives, not for decision-maker characteristics. For the effect of individual attribute (characteristics), it is referred to as the importance of an attribute (of a characteristic in reflecting the choice behaviour).

4.1. Data

The dataset contains travel mode choices of 1433 individuals in Australia who participated in an RP survey and a set of SP experiments. The participants were asked to provide information about their actual recent trips and about the same trips under 4 hypothetical scenarios detailing climate changes and mobility market changes related to self-driving vehicles and shared vehicles. More specifically, in the first stage of the experiment, sociodemographic information was collected form the respondents in the form of a questionnaire. Subsequently, the respondents were asked to provide the actual recent trip information such as the number of trips, modes, durations, etc. In the second stage, 4 hypothetical situations such as varying degree of mobility market changes, climate changes, etc. were given to the respondents, and the respondents were asked to choose a travel mode for each of the trips reported in the first stage. The travel modes available were private vehicles (“Private”), public transport (“Public”), walk or bike (“ActiveTravel”), hail and ride (“Hail”), and shared vehicles (“Share”). More detailed description on the experimental context can be found in Zhou et al. (2020).

Table 1 shows a list of the variables considered in the model. Sociodemographic variables for the participants include gender, age, income, education, number of cars, and number of two-wheelers. Indicator variables capture whether a participant provided or not gender or income information. The number of trips from the day before the interview was derived from the reported trip information, and then SP-specific variables included the percentage of shared trips, the percentage of self-driving vehicles and the level of climate change. Lastly, the indicator “RPInd” corresponds to the indicator D_{nt}^{RP} in eq. (8) identifying the choices made in the RP context.

4.2. Inferred preferences

The estimated model consists of the posterior samples of all the unknown parameters $\boldsymbol{\beta}$ in eq. (19). Most relevantly, the probability parameters p_{md} allow inferring the attendance/importance of each attribute m in the choice process, the mean parameters of the intercepts μ_{0di} and the mean parameters of the magnitude parameters μ_{mi} for each attribute m allow inferring the averaged effect of the intercept and each attribute among those individuals who have taken the attributes into consideration in the choice process or have the attributes as important determining factors, and the scale parameter λ allows inferring potential difference in choice uncertainty under different situations.

4.2.1. Level of attendance or importance of attributes

Table 2 shows the estimated medians and the 90% credible intervals for the probability parameters p_{md} of each attribute in the RP and SP data. The estimates in this table can be interpreted as the proportions of respondents who considered the corresponding

Table 1
Variables in the mode choice dataset.

Name	Description	Domain
Gender	indicator of females	{0, 1}
GenderDummy	indicator of respondents with unidentified gender	{0, 1}
Income	household income (17 categories)	{1, …, 17}
IncomeDummy	indicator of respondents with unknown income	{0, 1}
Age	age of the respondents (8 categories)	{1, …, 8}
Education	education of the respondents (6 categories)	{1, …, 6}
TotalCars	number of cars owned by the household	{0, …, 5}
TotalTW	number of two-wheelers owned by the household	{0, …, 5}
TripCount	number of trips of the respondents from the day before	{0, …, 30}
ShareTripPerc	percentage of shared trips (SP-specific, 4 categories)	{1, …, 4}
SelfDrivePerc	percentage of self-driving vehicles (SP-specific, 4 categories)	{1, …, 4}
ClimChange	level of climate change (SP-specific, 3 categories)	{1, …, 3}
RPInd	indicator of RP observations	{0, 1}
Intercept	constant	{1}

attributes in their choices or have the attributes as important determining factors.

The parameter estimates reveal that the probabilities of education as attributes having non-zero effect differ significantly in the RP and SP situations (0.203 vs. 0.499). Personal characteristics seem to be important in hypothetical situations far more than in actual situations, where situational constraints play a far more significant role. In fact, parameter estimates reveal that income has a non-zero effect in RP situations more likely than in SP ones (0.738 vs. 0.504). Not surprisingly, the income level is a constraint when choices are made in actual situations because of real financial cost considerations, whereas is considered less when choices are made in hypothetical situations because of the non-realisation of the financial costs. Also, parameter estimates reveal that age shows an even greater difference between the RP and SP situations (0.763 vs. 0.479). Not surprisingly, age implies physical limitations that manifest in actual situations but are ignored in hypothetical ones.

The parameter estimates also show that the estimated probabilities for "TotalCars", "TotalTW" and "TripCount" are very different between the RP and SP situations. The estimates for "TotalCars" and "TripCount" show higher importance in the RP situations (0.855 and 0.953) than in the SP situations (0.480 and 0.515). Possibly, the opportunity cost of owning cars, the time constraints and availability of public transport, the physical constraints of active travel, and the financial constraints for hail and ride, tend to be ignored more often in hypothetical situations. Interestingly, the estimates for "TotalTW" indicate a higher importance in the SP situations (0.010 vs. 0.559). Possibly, two-wheelers are used in Australia as a secondary mode because of a number of reasons (e.g., weather conditions, comfort level) that are downplayed or ignored in hypothetical situations.

The attributes that are specific to the SP experiments are all attended by a considerable portion of the respondents, and the proposed model allows to observe that the percentages of shared trips (0.698) is considered more often than the percentages of self-driving vehicles and the level of climate change (0.457 and 0.407). Moreover, the parameter estimates show that non-zero inertia effects exist for all the transport modes, with higher inertia for shared vehicles (0.708) and lower inertia for hail rides (0.511). The presence of SP-specific attributes likely explains the difference in attendance of non-zero intercepts of travel modes with respect to the RP part of the data (0.563vs. 0.967). When additional information is not recorded in the dataset, as in this case happens for the actual observations, the proposed model controls for potential bias that may appear in the parameter estimates.

4.2.2. Magnitude effect of attributes

Table 3 presents the population distributions of the magnitude parameters of the attributes and the SP scale parameter. The behavioural interpretation of the medians of the posterior samples is quite intuitive: private vehicles are more likely to be chosen by individuals with lower education level, higher number of cars, lower number of two-wheelers and lower number of trips, as well as by older females; public transport is more probable to be chosen by younger individuals with higher education and lower number of cars and trips; hail and shared rides are more likely to be chosen by individuals with lower number of trips. When looking at the SP-specific attributes, a higher percentage of shared trips is related to a higher probability of choosing shared rides, a higher percentage of self-driving vehicles is associated with a decrease in the probability of choosing private transport.

The magnitudes of the inertia parameters show that the higher level of inertia is for public transport (0.624) whereas the lower one is for hail and shared rides. Possibly, in hypothetical situations individuals would repeat sustainable transport choices more than others. Moreover, the presence of SP-specific attributes implies that the magnitude of the intercepts is lower for the SP with respect to the RP situations, also a pattern similar to the one of the respective indicators.

The estimated scale parameter is larger than one (2.153), which is an indication of a larger variance in the SP choices. This result is in line with the existing literature about choice making in hypothetical situations, but the estimate of the scale ratio is larger than the usual values observed in the existing literature (e.g., Hensher and Bradley 1993; Brownstone et al. 2000). However, the proposed

Table 2
Estimated distributions of the probabilities p_{md} of attribute attendance/importance in RP and SP situations (medians and 90% credible intervals).

	RP		SP	
	Median	90% interval	Median	90% interval
Gender	0.890	(0.790,0.984)	0.555	(0.490,0.623)
GenderDummy	0.223	(0.002,0.666)	0.521	(0.215,0.794)
Income	0.738	(0.566,0.896)	0.504	(0.459,0.548)
IncomeDummy	0.386	(0.065,0.724)	0.488	(0.412,0.566)
Age	0.763	(0.716,0.814)	0.479	(0.429,0.525)
Education	0.203	(0.111,0.301)	0.499	(0.463,0.537)
TotalCars	0.855	(0.823,0.884)	0.480	(0.429,0.524)
TotalTW	0.010	(0.001,0.191)	0.559	(0.491,0.631)
TripCount	0.953	(0.848,0.998)	0.515	(0.486,0.545)
InertiaDPrivate			0.548	(0.513,0.579)
InertiaDPublic			0.523	(0.487,0.556)
InertiaDActiveTravel			0.588	(0.539,0.633)
InertiaDHail			0.511	(0.322,0.712)
InertiaDShare			0.708	(0.385,0.891)
ShareTripPerc			0.698	(0.614,0.787)
SelfDrivePerc			0.457	(0.383,0.524)
ClimChange			0.407	(0.305,0.512)
Intercept	0.967	(0.936,0.985)	0.563	(0.508,0.624)

Table 3
Estimated parameters for the population distributions of the magnitude parameters μ_{0di}, μ_{mi} and λ (medians with 90% credible intervals in brackets).

	Private	Public	ActiveTravel (baseline)	Hail	Share
Gender	0.098 (0.024,0.176)	−0.001 (-0.085,0.086)	–	0.120 (-0.106,0.429)	−0.006 (-0.199,0.187)
GenderDummy	−0.240 (-0.635,0.237)	0.242 (-0.213,0.670)	–	−0.085 (-1.000,0.654)	−0.053 (-0.748,0.653)
Income	0.010 (-0.002,0.023)	0.008 (-0.006,0.023)	–	0.014 (-0.015,0.054)	−0.003 (-0.028,0.023)
IncomeDummy	0.048 (-0.104,0.208)	0.099 (-0.086,0.274)	–	0.165 (-0.222,0.664)	−0.049 (-0.343,0.247)
Age	0.042 (0.013,0.069)	−0.091 (-0.121,-0.060)	–	0.020 (-0.055,0.145)	0.009 (-0.055,0.070)
Education	−0.045 (-0.065,-0.025)	0.031 (0.008,0.052)	–	0.021 (-0.020,0.069)	0.002 (-0.034,0.037)
TotalCars	0.264 (0.223,0.310)	−0.152 (-0.200,-0.104)	–	−0.007 (-0.109,0.104)	0.035 (-0.049,0.119)
TotalTW	−0.110 (-0.165,-0.053)	0.026 (-0.031,0.086)	–	−0.024 (-0.153,0.077)	−0.013 (-0.099,0.082)
TripCount	−0.018 (-0.036,-0.000)	−0.066 (-0.086,-0.046)	–	−0.080 (-0.180,-0.020)	−0.061 (-0.122,-0.018)
InertiaDPrivate	0.589 (0.492,0.689)	–	–	–	–
InertiaDPublic	–	0.624 (0.540,0.714)	–	–	–
InertiaDActiveTravel	–	–	0.548 (0.475,0.621)	–	–
InertiaDHail	–	–	–	0.293 (0.124,0.475)	–
InertiaDShare	–	–	–	–	−0.197 (-0.480,0.039)
ShareTripPerc	−0.001 (-0.013,0.012)	−0.008 (-0.022,0.006)	–	−0.011 (-0.030,0.006)	0.018 (0.002,0.035)
SelfDrivePerc	−0.013 (-0.027,-0.001)	0.000 (-0.014,0.015)	–	0.001 (-0.018,0.019)	−0.001 (-0.019,0.014)
ClimChange	−0.000 (-0.018,0.016)	0.001 (-0.017,0.021)	–	−0.005 (-0.029,0.018)	−0.020 (-0.042,0.002)
ScaleSP	–	–	2.153 (1.897,2.464)	–	–
Intercept RP	1.564 (1.018,2.099)	2.418 (1.825,2.989)	–	−3.266 (-8.332,-1.216)	−1.535 (-3.039,-0.514)
Intercept SP	0.000 (-0.129,0.130)	−0.044 (-0.190,0.087)	–	−0.453 (-0.667,-0.275)	−0.228 (-0.388,-0.085)

Lastly, it is important to note that the estimated parameters in Table 3 are obtained conditional on the corresponding individual-specific indicators being one, namely they are estimated on a subset of observations. Accordingly, the behavioural interpretation should use a word of caution when commenting on modes where a subset is extracted from an already small number of observations.

model isolates the scale parameter from other behavioural discrepancies between the SP and RP data, such as ignoring situation constraints, non-attending attributes, and misinterpreting attributes. Hence, the large value of the estimated scale parameter underlines the importance of correcting for the biases associated with each of the behavioural discrepancies rather than combining most of them (with the exception of decision inertia) within the estimation of the scale parameter.

4.2.3. Effect of attributes

Having presented the indicator and the magnitude parameters, it is informative to explore their joint distribution in order to illustrate the effect of attributes at the individual level. Fig. 1 illustrates the joint distributions of the posterior samples of individual specific parameters associated with the key attributes. Generally, the joint distributions present a peak at zero, corresponding largely to the observations with indicator parameter equal to zero, and the values from the product of indicator and magnitude parameters, corresponding to the effects of the attributes. It should be noted that the distributions on the RP part of the observations are different from the ones on the SP part because of the scale parameter effect: the non-zero effect of any attribute is harder to be discerned from the zero in the noisier SP part of the observations.

Fig. 2 explores the joint distributions of the posterior samples of individual specific parameters ($I^d_{mn}\alpha^d_{mi}$) associated with the SP-specific attributes “ShareTripPerc”, “SelfDrivePerc” and “ClimChange”. The joint distributions mostly present a dispersion around the value zero, but an interpretation different from the previous one applies to this case. In fact, the attendance indicator for attribute “ShareTripPerc” is quite high (0.698) in comparing with those for the other two SP-specific attributes or for all SP attributes. Then, the joint distributions suggest that individuals have very different non-zero effect in terms of both the magnitude and the sign. This implies that there exists high heterogeneity across the individuals for the attribute “ShareTripPerc”, but in this case the heterogeneity is

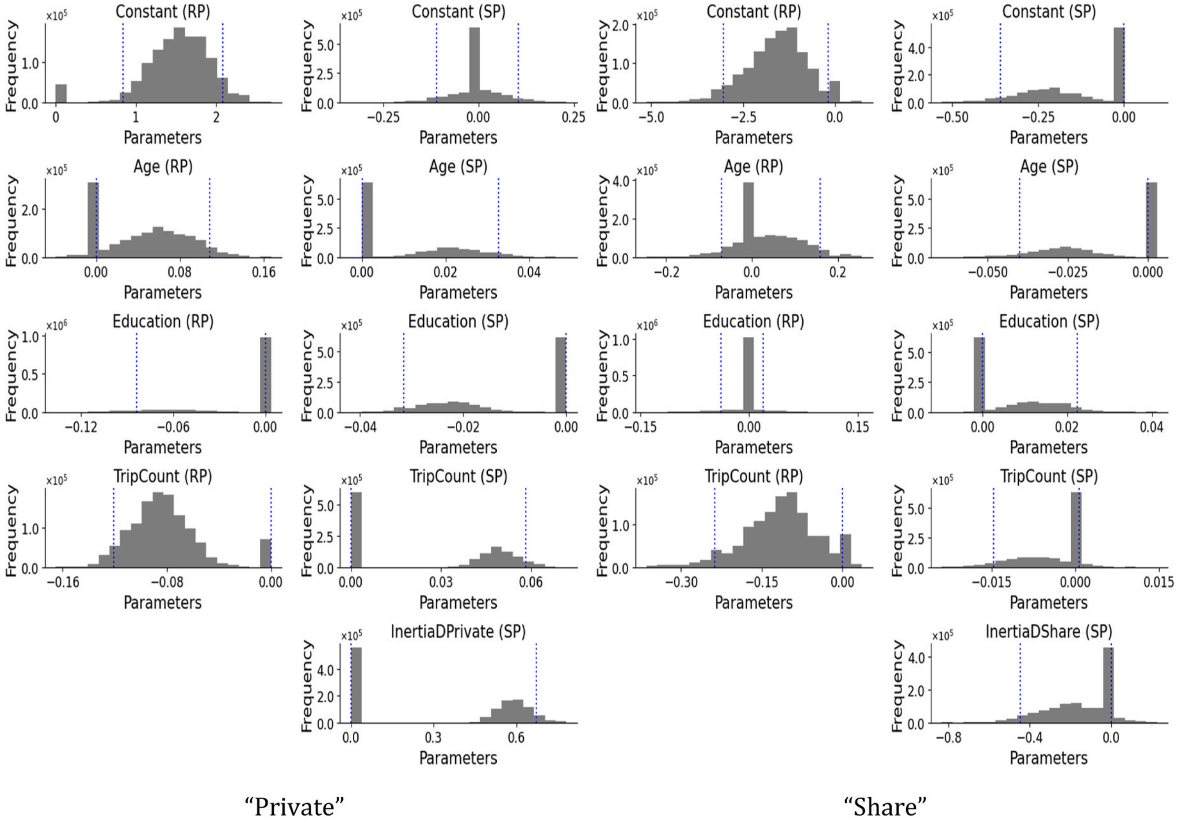

Fig. 1. Joint distributions of indicator and magnitude parameters ($I_{mn}^{d}\alpha_{mi}^{d}$) for selected attributes of travel modes "Private" and "Share" (90% credible intervals are indicated as dotted vertical lines).

actually because of different preferences for the travel modes rather than the result of non-attendance of attributes. It should be noted that this interpretation is possible because the proposed model has already accounted for inertia, scale difference, situation constraints, and attribute attendance.

From the behavioural perspective, the heterogeneity for the attribute "ShareTripPerc" can emerge from their possible misinterpretation. In the SP experiment, respondents were asked to choose while considering the three additional attributes "ShareTripPerc", "SelfDrivePerc" and "ClimChange". It is possible for example that the respondents had different knowledge about what shared trips are, different interpretation of the percentage as the whole market or the specific trip, different perception of how self-driving the vehicles will be, and different feel for rising sea levels and increasing average temperatures. The joint distributions in Fig. 2 reflect this misinterpretation that creates the heterogeneity in preferences (again, after having corrected for the other four behavioural discrepancies between the SP and RP contexts). Moreover, the joint distributions in Fig. 1 allow for a large proportion of indicator parameters being zero, which makes it easier to interpret attributes. This is the benefit of having carefully designed the prior structure and the parameter expansion in the proposed model, tapping into the potential of uncovering detailed information on the choice behaviour of the individuals without confounding heterogeneity from the attendance of attributes (which are easy to interpret) with heterogeneity from preferences for attributes (which are more difficult to interpret).

Having provided a visual representation of the joint distributions of the indicator and magnitude parameters, for completeness Table 4 shows the medians and 90% credible intervals for the joint distributions of the parameters associated to each attribute.

In line with what is observed in Figs. 1 and 2, the estimates show that zero belongs to most of the credible intervals, and the non-zero effects of attributes in the RP situation are more apparent than the ones in the SP situation. Most likely, these patterns arise naturally from the larger uncertainty in the SP experiments (as captured by the scale parameter). Moreover, the non-zero effects of the attributes are more evident for the first three modes than the remaining two. Most probably, these patterns arise from the selection of a subset of observations from an already limited number of observations for the less frequently chosen travel modes. Lastly, it should be noted that these estimates are different from the ones traditionally used for inference on the basis of a random parameter model. In fact, Table 4 reports the estimates for the product of one individual-specific parameter and one context-specific parameter rather than the traditional parameters of the population distributions. Accordingly, the estimates presented in Table 4 capture not only the uncertainty in the estimation of the population parameters, but also the one in the estimation of the individual-specific parameters and the context-specific parameters, which are at the root of the high level of heterogeneity observed in the joint distributions.

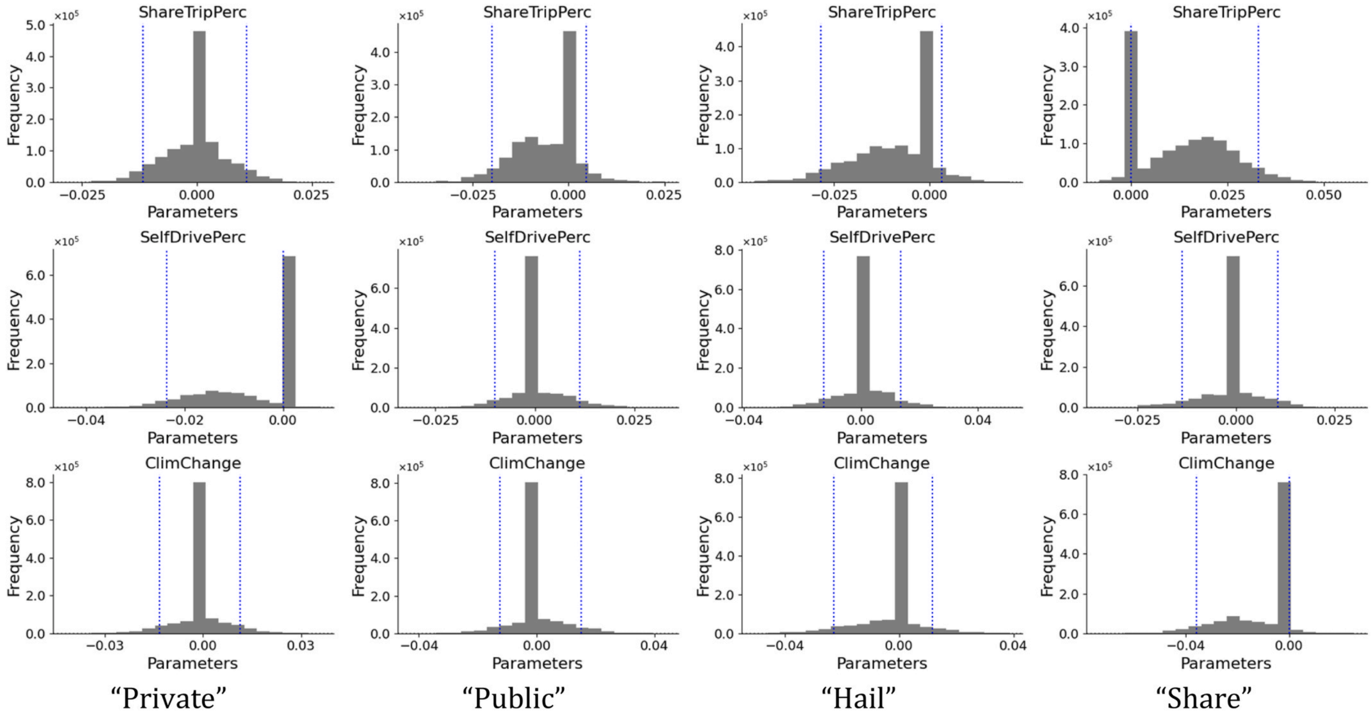

Fig. 2. Joint distributions of indicator and magnitude parameters ($I_{mn}^{d}\alpha_{mi}^{d}$) for the SP-specific attributes of the four travel modes (90% credible intervals are indicated as dotted vertical lines).

Table 4
Estimated parameters of the joint distributions of indicator and magnitude parameters $\boldsymbol{I}^{d}_{mn}\boldsymbol{\alpha}^{d}_{mi}$ (medians with 90% credible intervals in brackets).

	Private		Public		ActiveTravel (baseline)		Hail		Share	
	RP	SP	RP	SP	RP	SP	RP	SP	RP	SP
Gender	0.189 (0.000,0.359)	0.000 (-0.031,0.019)	−0.018 (-0.191,0.125)	0.003 (0.000,0.061)	–	–	0.217 (-0.179,0.841)	−0.001 (-0.068,0.000)	0.000 (-0.377,0.368)	0.000 (-0.039,0.023)
GenderDummy	0.000 (-0.561,0.389)	−0.041 (-0.442,0.000)	0.000 (0.000,1.300)	−0.029 (-0.518,0.000)	–	–	0.000 (-0.994,0.767)	0.000 (-0.306,0.072)	0.000 (-0.710,0.902)	0.000 (-0.330,0.037)
Income	0.000 (-0.015,0.030)	0.006 (0.000,0.019)	0.000 (-0.020,0.032)	0.002 (0.000,0.016)	–	–	0.012 (-0.022,0.104)	0.000 (-0.007,0.004)	0.000 (-0.051,0.043)	0.000 (-0.008,0.004)
IncomeDummy	0.000 (-0.335,0.083)	0.000 (0.000,0.289)	0.000 (-0.080,0.403)	0.000 (-0.024,0.097)	–	–	0.000 (-0.163,1.040)	0.000 (-0.090,0.064)	0.000 (-0.518,0.312)	0.000 (-0.085,0.061)
Age	0.048 (0.000,0.108)	0.000 (0.000,0.033)	−0.177 (-0.247,0.000)	0.000 (-0.000,0.021)	–	–	0.007 (-0.096,0.270)	0.000 (-0.017,0.008)	0.012 (-0.071,0.159)	0.000 (-0.040,0.000)
Education	0.000 (-0.084,0.000)	0.000 (-0.031,0.000)	0.000 (0.000,0.068)	0.000 (0.000,0.020)	–	–	0.000 (-0.005,0.069)	0.000 (0.000,0.026)	0.000 (-0.039,0.019)	0.000 (0.000,0.022)
TotalCars	0.505 (0.000,0.603)	0.000 (-0.001,0.026)	−0.241 (-0.342,0.000)	0.000 (-0.068,0.000)	–	–	−0.010 (-0.230,0.171)	0.000 (0.000,0.042)	0.037 (-0.103,0.216)	0.000 (-0.003,0.030)
TotalTW	0.000 (-0.000,0.000)	−0.121 (-0.174,0.000)	0.000 (0.000,0.000)	0.000 (-0.038,0.008)	–	–	0.000 (0.000,0.000)	0.000 (-0.036,0.020)	0.000 (0.000,0.000)	0.000 (-0.035,0.017)
TripCount	−0.084 (-0.121,0.000)	0.039 (0.000,0.058)	−0.105 (-0.147,0.000)	−0.015 (-0.033,0.000)	–	–	−0.146 (-0.347,0.000)	0.000 (-0.020,0.000)	−0.111 (-0.237,0.000)	0.000 (-0.015,0.001)
InertiaDPrivate	–	0.510 (0.000,0.673)	–	–	–	–	–	–	–	–
InertiaDPublic	–	–	–	0.538 (0.000,0.697)	–	–	–	–	–	–
InertiaDActiveTravel	–	–	–	–	–	0.500 (0.000,0.609)	–	–	–	–
InertiaDHail	–	–	–	–	–	–	–	0.101 (0.000,0.437)	–	–
InertiaDShare	–	–	–	–	–	–	–	–	–	−0.102 (-0.448,0.002)
ShareTripPerc	–	0.000 (-0.012,0.011)	–	−0.003 (-0.020,0.005)	–	–	–	−0.005 (-0.028,0.003)	–	0.013 (0.000,0.033)
SelfDrivePerc	–	0.000 (-0.024,0.000)	–	0.000 (-0.010,0.011)	–	–	–	0.000 (-0.013,0.014)	–	0.000 (-0.014,0.011)
ClimChange	–	0.000 (-0.013,0.011)	–	0.000 (-0.013,0.015)	–	–	–	0.000 (-0.023,0.011)	–	0.000 (-0.036,0.000)
Constant	1.546 (0.843,2.085)	0.000 (-0.111,0.104)	2.395 (1.650,2.981)	0.000 (-0.168,0.067)	–	–	−3.208 (-8.330,-0.751)	−0.323 (-0.628,0.000)	−1.511 (-3.031,-0.186)	−0.121 (-0.358,0.000)

4.3. Discussion

Given the presentation of the parameter estimates in the tables and figures in this section, it appears relevant to ask which type of parameter estimates should be preferred when making inference on the basis of joint RP and SP data.

The immediate answer is that the parameters to be preferred would be the population parameters presented in Table 3 because they are close to the ones used for inference with random parameter models. However, a more accurate answer is that it is important to know, at least approximately, how many observations played a role in the estimation of the population parameters presented in Table 3. To this extent, Table 2 provides a picture of the level of attendance/importance of each attribute and Table 4 goes full circle by introducing the joint distributions of indicator and magnitude parameters related to each attribute.

In light of the model formulation, it appears also relevant to ask whether the proposed model would complicate the interpretation of the parameter estimates or the calculation of some useful quantity such as the marginal rate of substitution. One might argue that there is a disadvantage in missing a general standard such as the traditionally used significance level. One might also argue that there are more parameters to comment upon and there are more distributions containing zero, and then perhaps question whether the means of the population distributions are significantly different from zero. However, the means of the population distributions at zero have a behavioural interpretation in the proposed model, and the advantages of the proposed modelling framework are substantial in terms of behavioural interpretation when considering the bias corrections that the proposed model is able to accommodate. Most notably, these corrections allow the proposed model to disentangle the impact caused by decision inertia for each alternative, scale parameter for the SP situations, level of attendance for each attribute and each choice situation, situation constraints, and possible misinterpretation of attributes.

For example, the model has a substantial advantage in disentangling the heterogeneity because of the level of attendance/importance of attributes (as in Fig. 1) from the actual heterogeneity because of different preference structures across individuals (as in Fig. 2). Traditionally, heterogeneity would be observed and modelled without reflecting about its causes, and possibly non-significance of the mean of the random parameter distribution would be found. In fact, traditionally the mean of the random parameters provides information about the average effect of the parameter over the entire sample. Instead, the distributions of the population parameters differentiate the percentage of the sample with non-zero effects from the magnitude of the effect itself, giving a more informative interpretation of the effects of attributes while looking at the choice process at the individual level. Thus, the proposed model is more informative and sensible by considering an attribute only influencing the choice behaviour of a proportion of respondents in a large sample rather than assuming that there is an average effect of the attribute for all the sample.

In light of the model formulation, it is natural to ask whether the proposed model would have a good potential for forecasting. Traditionally, forecasting would respond to a change in the value of attributes with the calculation or the choice-based simulation of average probabilities for all the alternatives. Implicitly, the assumption would be that the change in the value of attributes would occur for all the sample, including the individuals who might have ignored some attributes, with consequent bias in the prediction of the change in probabilities. The proposed model overcomes this issue, since the non-attendance of attributes for each individual is taken into account, and hence the change of an attribute would be considered only for the individuals that actually attend the same attribute. As a result, the predicted probabilities may be more accurate and in line with what the individuals actually considered in their choice behaviour. As mentioned previously, it should be noted that the introduction of indicator parameters corresponds to a variable selection approach that is commonly used in the forecasting literature (Smith 2000; Jochmann et al. 2010).

5. Conclusions

This study recognised that RP and SP data are two fundamental sources of information for preference studies across various disciplines and, owing to the distinctive advantages of each data type and their complementary nature, there is a substantial interest in making joint inferences from the combination of these two types of data. This allows one to not only combine the advantages of each type of data, but also make better inferences on the basis of all available information.

This study also recognised that there exist five behavioural discrepancies between the RP and SP data (decision inertia, scale difference, ignorance of situation constraints, non-attendance of attributes, and misinterpretation of attributes), and that the existing literature has traditionally tackled a few explicitly (i.e., decision inertia), and the others implicitly with the estimation of scale parameters. The main contribution of this study lies in disentangling these five behavioural discrepancies in an integrated modelling framework. The proposed model is inspired by the principles of variable selection and parameter expansion to correct for the biases introduced by these behavioural discrepancies, and are accommodated in a hierarchical Bayesian framework.

The design of the model formulation, the proposition of the hierarchical prior structure for model estimation, and the application to a travel mode choice case study illustrate the model is capable of not only accounting for the biases introduced by the aforementioned behavioural discrepancies, but also providing insights into parameter interpretation. Specifically, the traditional way of interpreting parameter estimates is substituted by a more informative way where the level of attendance of an attribute is disentangled from the measurement of the effect of the same attribute. The use of individual-specific indicators allows obtaining population distributions of parameter estimates that suggest the importance of knowing which attributes individuals consider in their choice behaviour, in combination with the well-established importance of knowing how much individuals consider those attributes. The discussion suggests that the model is more informative, again because of the features that allow disentangling attendance and magnitude of parameters. The most interesting result is the differentiation of the nature of heterogeneity as originated by either non-attending attributes or having different preference structures.

The proposed model might be argued to be elaborate and computationally demanding. Obviously, there does not exist a universally superior model as it generally depends on the research question, time constraints, etc. However, there exists a balanced model able to provide informative answers to behavioural questions. The Bayesian hierarchical framework for the proposed model provides a flexible way of modelling RP and SP data for making joint inferences and solving behavioural issues that were recognised long ago (Morikawa et al. 2002) but only partially solved. Future research could look at the potential forecasting capability of the proposed model and the calculation of useful quantities such as marginal rate of substitution, the possibility to relax the assumption that the attendance indicators have the same value across the alternatives, which would make it possible to further disentangle the issues with misinterpretation of attributes, as well as the assumption that the scale parameter has the same value across the observations. Of course, further elaboration of the model should consider research question and data availability to have a balanced and flexible model for the problem at hand. At last, a detailed comparison between the proposed model and similar alternatives could also provide valuable information on the potential ways for future improvements. But in such comparison, the use of likelihood value may not be informative due to the difference in model complexity. A more meaningful approach may be comparing the different information and inference that can be drawn from each alternative.

Author statement

Zili LI: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Software; Validation; Writing - original draft.

Simon P. WASHINGTON: Conceptualization; Formal analysis; Funding acquisition; Investigation; Methodology; Supervision; Validation; Writing - review & editing.

Zuduo Zheng: Conceptualization; Formal analysis; Funding acquisition; Investigation; Methodology; Supervision; Validation; Writing - review & editing.

Carlo G. PRATO: Formal analysis; Funding acquisition; Investigation; Methodology; Supervision; Validation; Writing - review & editing.

Declaration of competing interest

There is no any conflict of interest.

References

Adamowicz, Wiktor, Swait, Joffre, Boxall, Peter, Jordan, Louviere, Williams, Michael, 1997. Perceptions versus objective measures of environmental quality in combined revealed and stated preference models of environmental valuation. J. Environ. Econ. Manag. 32 (1), 65–84.

Allenby, Greg M., Lenk, Peter J., 1994. Modeling household purchase behavior with logistic normal regression. J. Am. Stat. Assoc. 89 (428), 1218–1231.

Andersson, Henrik, Arne, Risa Hole, Svensson, Mikael, 2016. Valuation of small and multiple health risks: a critical analysis of SP data applied to food and water safety. J. Environ. Econ. Manag. 75 (January), 41–53.

Andrews, Rick L., Srinivasan, T.C., 1995. Studying consideration effects in empirical choice models using scanner panel data. J. Market. Res. 32 (1), 30–41.

Baanders, Arianne N., 2002. Labour participation of the chronically ill: a profile sketch. Eur. J. Publ. Health 12 (2), 124–130.

Ben-Akiva, Moshe, Morikawa, Takayuki, 1990. Estimation of travel demand models from multiple data sources. In: International Symposium on Transportation and Traffic Theory, 11th, pp. 461–476 (Yokohama, Japan).

Ben-Akiva, M., Bradley, M., Morikawa, T., Benjamin, J., Novak, T., Oppewal, H., Rao, V., 1994. Combining revealed and stated preferences data. Market. Lett. 5 (4), 335–349.

Ben-Akiva, Moshe, Daniel, McFadden, Train, Kenneth, 2019. Foundations of stated preference elicitation: consumer behavior and choice-based conjoint analysis. Foundations and Trends® in Econometrics 10 (1–2), 1–144.

Bhat, Chandra R., 1997. Work travel mode choice and number of non-work commute stops. Transp. Res. Part B Methodol. 31 (1), 41–54.

Bhat, Chandra R., Castelar, Saul, 2002. A unified mixed logit framework for modeling revealed and stated preferences: formulation and application to congestion pricing analysis in the san francisco bay area. Transp. Res. Part B Methodol. 36 (7), 593–616.

Börjesson, Maria, 2008. Joint RP–SP data in a mixed logit analysis of trip timing decisions. Transport. Res. E Logist. Transport. Rev. 44 (6), 1025–1038.

Bradley, Mark A., Daly, Andrew J., 1997. Estimation of logit choice models using mixed stated preference and revealed preference information. In: Understanding Travel Behaviour in an Era of Change, pp. 209–232.

Broadway, Barbara, Kalb, Guyonne, Li, Jinhu, Scott, Anthony, 2017. Do financial incentives influence GPs' decisions to do after-hours work? A discrete choice labour supply model. Health Econ. 26 (12), e52–e66.

Brownstone, David, Bunch, David S., Kenneth, Train, 2000. Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles. Transp. Res. Part B Methodol. 34 (5), 315–338.

Cherchi, Elisabetta, Juan de Dios Ortúzar, 2006. On fitting mode specific constants in the presence of new options in RP/SP models. Transport. Res. Pol. Pract. 40 (1), 1–18.

Cherchi, Elisabetta, Juan de Dios Ortúzar, 2011. On the use of mixed RP/SP models in prediction: accounting for systematic and random taste heterogeneity. Transport. Sci. 45 (1), 98–108.

Cobb-Walgren, Cathy, J., Ruble, Cynthia A., Donthu, Naveen, 1995. Brand equity, brand preference, and purchase intent. J. Advert. 24 (3), 25–40.

Erdem, Tülin, Keane, Michael P., 1996. Decision-making under uncertainty: capturing dynamic brand choice processes in turbulent consumer goods markets. Market. Sci. 15 (1), 1–20.

Garcia, Andrea N., Kuo, Tony, Arangua, Lisa, Pérez-Stable, Eliseo J., 2018. Factors associated with medical school graduates' intention to work with underserved populations. Acad. Med. 93 (1), 82–89.

Gelfand, Alan E., Smith, Adrian F.M., 1990. Sampling-based approaches to calculating marginal densities. J. Am. Stat. Assoc. 85 (410), 398–409.

Gelman, A., 2006. Prior distributions for variance parameters in hierarchical models. Bayesian Analysis 1, 515–533.

Gelman, Andrew, Carlin, John B., Stern, Hal S., Dunson, David B., Aki, Vehtari, Rubin, Donald B., 2013. Bayesian Data Analysis, third ed. Chapman and Hall/CRC. Bayesian Data Analysis, Third Edition.

Gilbride, Timothy J., Allenby, Greg M., 2004. A choice model with conjunctive, disjunctive, and compensatory screening rules. Market. Sci. 23 (3), 391–406.

Guadagni, Peter M., Little, John D.C., 1983. A logit model of brand choice calibrated on scanner data. Market. Sci. 2 (3), 203–238.

Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57 (1), 97–109.

Hensher, David A., Bradley, Mark, 1993. Using stated response choice data to enrich revealed preference discrete choice models. Market. Lett. 4 (2), 139–151.

Hensher, David A., Rose, John M., Greene, William H., 2005. Applied Choice Analysis. Applied Choice Analysis: A Primer. Cambridge University Press, Cambridge.
Hensher, David A., Rose, John M., Greene, William H., 2008. Combining RP and SP data: biases in using the nested logit 'trick' – contrasts with flexible mixed logit incorporating panel and scale effects. J. Transport Geogr. 16 (2), 126–133.
Huang, Alan, Wand, M.P., 2013. Simple marginally noninformative prior distributions for covariance matrices. Bayesian Analysis 8 (2), 439–452.
Jochmann, Markus, Koop, Gary, Strachan, Rodney W., 2010. Bayesian forecasting using stochastic search variable selection in a VAR subject to breaks. Int. J. Forecast. 26 (2), 326–347.
Kuo, Lynn, Mallick, Bani, 1998. Variable selection for regression models. Sankhya: The Indian Journal of Statistics, Series B 60, 65–81.
Louviere, Jordan J., Hensher, David A., Swait, Joffre D., Adamowicz, Wiktor, 2000. Combining Sources of Preference Data, vols. 227–51. Cambridge University Press, Cambridge. Stated Choice Methods.
McFadden, Daniel, 1974. Conditional logit analysis of qualitative choice behaviour. In: Zarembka, P. (Ed.), Frontiers in Econometrics, vols. 105–42. Academic Press, New York.
Mentzakis, Emmanouil, Stefanowska, Patricia, Hurley, Jeremiah, 2011. A discrete choice experiment investigating preferences for funding drugs used to treat orphan diseases: an exploratory study. Health Econ. Pol. Law 6 (3), 405–433.
Miller, Eric J., Roorda, Matthew J., Carrasco, Juan Antonio, 2005. A tour-based model of travel mode choice. Transportation 32 (4), 399–422.
Morikawa, Taka, Ben-Akiva, Moshe, McFadden, Daniel, 2002. Discrete choice models incorporating revealed preferences and psychometric data. Adv. Econom. 29–55.
O'Hara, R.B., Sillanpää, M.J., 2009. A review of bayesian variable selection methods: what, how and which. Bayesian Analysis 4 (1), 85–117.
Shin, Sangwoo, Misra, Sanjog, Horsky, Dan, 2012. Disentangling preferences and learning in brand choice models. Market. Sci. 31 (1), 115–137.
Sillanpää, Mikko J., Bhattacharjee, Madhuchhanda, 2005. Bayesian association-based fine mapping in small chromosomal segments. Genetics 169 (1), 427–439.
Smith, Michael, 2000. Modeling and short-term forecasting of new south wales electricity system load. J. Bus. Econ. Stat. 18 (4), 465–478.
Smith, Michael, Kohn, Robert, 2002. Parsimonious covariance matrix estimation for longitudinal data. J. Am. Stat. Assoc. 97 (460), 1141–1153.
Vij, Akshay, Carrel, André, Joan, L., Walker, 2013. Incorporating the influence of latent modal preferences on travel mode choice behavior. Transport. Res. Pol. Pract. 54 (August), 164–178.
Whitehead, John C., Pattanayak, Subhrendu K., Van Houtven, George L., Gelso, Brett R., 2008. Combining revealed and stated preference data to estimate the nonmarket value of ecological services: an assessment of the state of the science. J. Econ. Surv. 22 (5), 872–908.
Ye, Runing, Titheridge, Helena, 2017. Satisfaction with the commute: the role of travel mode choice, built environment and attitudes. Transport. Res. Transport Environ. 52 (May), 535–547.
Zhou, Fan, Zheng, Zuduo, Whitehead, Jake, Washington, Simon, Perrons, Robert K., Page, Lionel, 2020. Preference heterogeneity in mode choice for car-sharing and shared automated vehicles. Transport. Res. Pol. Pract. 132, 633–650.