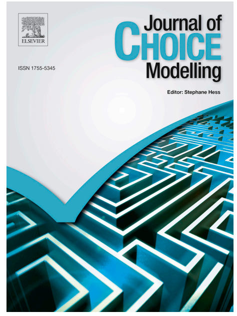
Journal Pre-proof



Quantifying behavioural difference in latent class models to assess empirical identifiability: Analytical development and application to multiple heuristics

Felipe Gonzalez-Valdes, Benjamin G. Heydecker, Juan De Dios Ortúzar

PII: S1755-5345(22)00014-8

DOI: https://doi.org/10.1016/j.jocm.2022.100356

Reference: JOCM 100356

To appear in: *Journal of Choice Modelling*

Received Date: 5 July 2021

Revised Date: 21 February 2022

Accepted Date: 30 March 2022

Quantifying behavioural difference in latent class models to assess empirical identifiability: analytical development and application to multiple heuristics

GONZALEZ-VALDES, Felipe[1]; HEYDECKER, Benjamin G.[2]; ORTÚZAR, Juan de Dios[3]

ABSTRACT

Latent class (LC) models have been used for decades. In some cases, models of this kind have exhibited difficulties in identifying distinct classes. Identifiability is key to determining the presence or absence of the different population cohorts represented by the latent classes. Theoretical identifiability addresses this issue in general, but no empirical identifiability analysis of this kind of model has been performed previously. Here, we analyse the theoretical properties of LC models to establish necessary conditions on the classes to be identifiable jointly. We then, establish a measure of behavioural difference and relate it to empirical identifiability; this measure highlights factors that are crucial for identifiability. We show how these factors affect identifiability through simulation experiments in which classes are known, and test elements such as the proportion of individuals belonging to each latent class, different correlation structures and sample sizes. In our experiments, each choice heuristic belongs to a distinct latent class. We present a graphical diagnostic that supports the measure of behavioural difference that promotes identifiability and provide examples of model non-identifiability, partial identifiability, and strong identifiability. We conclude by discussing how non-identifiability can be detected and understood in ways that will inform survey design and analysis.

Keywords: latent classes, empirical identifiability, discrete choice modelling

---

1 Department of Transport Engineering and Logistics, Pontificia Universidad Católica de Chile, Vicuña Mackenna 4860, Macul, Santiago, Chile; Tel: +56-2-2354 4270; e-mail: fagonzalezv@uc.cl

2 Centre for Transport Studies, University College London, Gower Street, London WC1E 6BT, England; Tel: +44-20-7679 1553; e-mail: b.heydecker@ucl.ac.uk

3 Department of Transport Engineering and Logistics, Institute in Complex Engineering Systems, BRT + Centre of Excellence, Pontificia Universidad Católica de Chile, Vicuña Mackenna 4860, Macul, Santiago, Chile; Tel: +56-2-2354 4822; e-mail: jos@ing.puc.cl

1. INTRODUCTION

Latent class (LC) models can be used to represent finite mixtures of distinct groups of individuals (Kamakura and Russell, 1989). They have been widely applied in recent decades either exclusively with exogenous variables (Swait and Adamowicz, 2001; Rossetti et al., 2018) or in conjunction with latent variables in a MIMIC model (Huang and Bandeen-Roche, 2004; Hess and Stathopoulos, 2013), with diffuse choice sets (Ben-Akiva and Boccara, 1995), and either using only utility maximisation heuristics or adopting a different choice heuristic for each latent class (Hess et al., 2012; Gonzalez-Valdes and Raveau, 2018).

A key issue concerning LC models is their identifiability, which is related to the possibility of drawing inference from observed samples about an underlying theoretical structure that is observationally unique[1]. Rothenberg (1971) examined the identifiability of parametric models, concluding that this required the information matrix to be non-singular. Walker and Ben-Akiva (2002) investigated theoretical and empirical identifiability. Here, we focus on the latter, where the model theoretically can be identified, but due to the data and model structure, the Hessian matrix is singular or nearly so (Chiou and Walker, 2007; Cherchi and Ortúzar, 2008), leading to poor estimates of model parameters and impeding empirical identification.

In LC models, identifiability informs about distinguishing different behaviour types and estimating the parameters that govern them, with the behaviour of each individual in the population being described as a linear combination of the theoretical constructs[2]. The identifiability of LC models has been studied to varying extents. Huang and Bandeen-Roche (2004) explored theoretical identifiability in LC models specifying conditions of the components of a latent class – latent variable choice model required to achieve it. However, requirements for empirical identifiability

[1] Identifiability of a model is achieved when no other model is observationally equivalent and no different set of parameters yield the same result. Drawing inferences from observed samples depend on the estimation method: under maximum likelihood, identifiability is a condition of the estimation whereas under the Bayesian approach it is a feature that can be assessed post-estimation. Nonetheless, the concept of identifiability is independent of the estimation method.

[2] One interpretation of the theoretical construct of latent classes is that they represent individuals according to similarities in their behaviour, although another is that the classes represent groups of the individuals themselves. We adopt the former interpretation even when there is a continuum of behaviours, in which case we use classes to represent clusters of them. However, for the sake of simplicity, here we develop our analysis according to the latter interpretation.

of models that have no latent variables have not been addressed thoroughly. Thus, this paper focuses on determining conditions necessary for empirical identifiability in the absence of latent variables.

Among the applications of LC models, the one that motivated the present study is when multiple choice heuristics are considered. Success has been reported in the literature for LC models under a single heuristic with multiple parameter sets (e.g. Greene and Hensher, 2003), but few have successfully presented identifiable multiple heuristic models. Indeed, these LC models have resorted to latent variables (Hess and Stathopoulos, 2013) and normalisations (Leong and Hensher, 2012) for identifiability. Here, after establishing analytical conditions for identifiability, we show how they apply in practice to the challenge of identifying multiple heuristics.

Connecting both of these objectives, this paper investigates the empirical identifiability of LC models when only exogenous variables are used (i.e. without latent variables). To understand this, we first develop a theoretical framework to analyse the interaction of the governing forces of identifiability and show that the ratio of class-conditional probability to the model-wide probability of observed choices is crucial. Then, we investigate the use of this framework by conducting a battery of Monte Carlo simulation experiments in a realistic transport context. In this, we follow the approach proposed by Chiou and Walker (2007) to explore influences on identifiability. The simulation of latent classes is performed in the context of multiple-choice heuristics to investigate identifiability. Each of three distinct choice heuristics is tested against a linear random utility maximisation (RUM) model to assess the identifiability of that combination. We explore drivers for non-identifiability that are exemplified by the scenario of multiple-choice heuristics. Finally, we show how the results of this study provide a framework for practitioners to design surveys and experiments of LC models.

The remainder of this paper is organised as follows. Section 2 develops a theoretical framework to investigate empirical identifiability and provides a metric to explain the reasons for non-identifiability. Section 3 describes the specification and execution of a battery of empirical experiments. Section 4 analyses the results of the experiments and relates them to the drivers of identifiability within the theoretical framework; indeed, this section is helpful for practitioners to understand possible reasons for lack of identifiability. It also shows how to connect reasons beyond those described here to the overarching theoretical framework discussed in Section 2. Finally, Section 5 concludes the paper summarising the main findings and tools.

2. ANALYTICAL DEVELOPMENT

We develop a theoretical framework based on maximum likelihood estimation that facilitates understanding of the identifiability of LC models. We first analyse a binary case in which the simple structure illuminates the underlying phenomena. Then, we generalise this to the case of multi-classes. In each analysis, we establish the first-order optimality conditions on the likelihood function to understand when coexisting classes can be identifiable, which we refer to as *theoretical identifiability*. Finally, the Hessian matrix of the likelihood function is analysed to relate identifiability to features of the model.

The results of applying this framework can be assessed according to the definition of identifiability introduced by Gu and Xu (2020). Thus, *strict identifiability* is achieved when all parameters of the model are recovered accurately. *Partial identifiability* is achieved when a range of parameter values yield similar model performance. Finally, *non-identifiability* arises when estimation results in a single class.

2.1 Binary Case

2.1.1 Latent classes with constant class membership function

Suppose that individuals align their behaviour to one of two latent classes, denoted as a and b, with probabilities π_a and $\pi_b = (1 - \pi_a)$ respectively. Let $P_{cqi}(\theta)$ be the probability that according to class $c \in \{a, b\}$ with parameters θ, individual q chooses alternative i. Then, $P_{qi}(\theta, \pi_a)$, the probability of individual q choosing alternative i under the LC model, is given by (1):

$$P_{qi}(\theta, \pi_a) = \pi_a P_{aqi}(\theta) + (1 - \pi_a) P_{bqi}(\theta) \tag{1}$$

The log-likelihood function of this model is given by (2), where $P_{cq*}(\theta)$ represents the probability that individual q would have chosen their selected alternative aligning their behaviour to latent class c:

$$l(\theta, \pi_a) = \sum_q ln\left(\pi_a\ P_{aq*}(\theta) + (1 - \pi_a)\ P_{bq*}(\theta)\right). \tag{2}$$

The maximum value of this likelihood function could arise either at a boundary or at an interior value of π_a. In the case of a boundary solution (i.e. $\pi_a \in \{0,1\}$), the optimal model consists of a single latent class: a when $\pi_a = 1$, or b when $\pi_a = 0$. By contrast, in the case of an interior solution (i.e. $\pi_a \in (0, 1)$), the two classes of individuals coexist in a mixture model corresponding to simultaneous identification of the two distinct latent classes. Thus, when an interior solution arises, it reflects theoretical identifiability[3].

The solution (interior or boundary) depends upon the losses and gains in likelihood associated with including an additional class in the model and, therefore, reducing the proportion of the complementary one. Class *a* may perform better than class *b* for some observations, with the reverse occurring for other observations. Including a second class, *b*, would improve the likelihood for the latter observations. However, in cases where the first class *a* performs better, there would be a loss of likelihood due to the reduction of its proportion in the model. The balance between these two changes in performance determines the type of solution obtained (i.e. whether the solution is a boundary or an interior one). A boundary solution will be obtained when it is optimal for the model to consider a single class of individuals, corresponding to the case where the improvement in likelihood from the inclusion of a second class does not compensate for the associated losses. Some examples illustrating these cases are shown in Appendix A.

In the case of an interior solution when identifiability of the class membership component is possible, likelihood is maximised when the likelihood function is stationary with respect to variations in the class membership probability π_a. This can be detected as an interior point at which the derivative of the log-likelihood function equals zero. Among the variables to examine, an interesting one is precisely π_a, because it indicates the proportions of the two classes and, therefore, connects them in the model. This first-order stationarity condition regarding π_a is analysed next.

[3] This condition is related to theoretically identification only of the class membership component, which is the focus of this paper.

We start by considering the case where the class membership function π_a is constant across the population (i.e. the probability of class membership is the same for every individual). For the context of multiple-choice heuristics that we explore later, this is the most frequent formulation (Adamowicz & Swait, 2013; Araña et al., 2008; Balbontin et al., 2017; Hess et al., 2012; McNair et al., 2012). Under this specification, the following theorem describes the optimality of estimation that corresponds to the coexistence of two latent classes:

THEOREM 1*: Two latent classes coexist optimally in a discrete choice model with constant class membership function if the vector θ of estimated parameters satisfies the balance specified by (3):*

$$\sum_q \frac{P_{aq*}(\theta)}{P_{q*}(\theta)} = \sum_q \frac{P_{bq*}(\theta)}{P_{q*}(\theta)} \tag{3}$$

where $P_{q*}(\theta, \pi_a) = \pi_a\, P_{aq*}(\theta) + (1 - \pi_a)\, P_{bq*}(\theta)$ denotes the modelled probability that individual q chooses the chosen alternative (consistent with (2)).

PROOF: For an interior solution, the first-order condition for the maximisation is given by (4):

$$\begin{gathered} \frac{\partial l(\theta, \pi_a)}{\partial \pi_a} = 0 \\ \Leftrightarrow \sum_q \frac{P_{aq*}(\theta) - P_{bq*}(\theta)}{\pi_a\, P_{aq*}(\theta) + (1 - \pi_a)\, P_{bq*}(\theta)} = 0\,. \end{gathered} \tag{4}$$

Manipulation of (4) leads to (5):

$$\sum_q \frac{P_{aq*}(\theta)}{\pi_a\, P_{aq*}(\theta) + (1 - \pi_a)\, P_{bq*}(\theta)} = \sum_q \frac{P_{bq*}(\theta)}{\pi_a\, P_{aq*}(\theta) + (1 - \pi_a)\, P_{bq*}(\theta)} \tag{5}$$

Using the definition of $P_{q*}(\theta)$, this is equivalent to (3).

Equations (3) and (5) show that a balance is achieved when it is optimal for the model to include both latent classes. This balance is given by the sum of the ratio of the likelihoods of the class to the complete model. This expression quantifies the balance dynamics of gains and losses in the likelihood function associated with the introduction of a second latent class to the model.

The magnitude of this sum is described by Theorem 2:

THEOREM 2*: Two latent classes coexist optimally in a discrete choice model with constant class membership function if the balance quantity in (3) is equal to the sample size Q.*

PROOF: Expanding the left-hand side of (3) leads to (6):

$$\sum_q \frac{P_{aq*}(\theta)}{P_{q*}(\theta,\pi_a)} = \sum_q \frac{\pi_a\, P_{aq*}(\theta)}{P_{q*}(\theta,\pi_a)} + \sum_q \frac{(1-\pi_a)\, P_{aq*}(\theta)}{P_{q*}(\theta,\pi_a)}$$

$$= \sum_q \frac{\pi_a\, P_{aq*}(\theta)}{P_{q*}(\theta,\pi_a)} + \sum_q \frac{(1-\pi_a)\, P_{aq*}(\theta)}{P_{q*}(\theta,\pi_a)} + \sum_q \frac{(1-\pi_a)\, P_{bq*}(\theta) - (1-\pi_a) P_{bq*}(\theta)}{P_{q*}(\theta,\pi_a)}$$

$$\Rightarrow \sum_q \frac{P_{aq*}(\theta)}{P_{q*}(\theta,\pi_a)} = \sum_q \frac{\pi_a\, P_{aq*}(\theta) + (1-\pi_a)\, P_{bq*}(\theta)}{P_{q*}(\theta,\pi_a)} + (1-\pi_a) \sum_q \frac{P_{aq*}(\theta) - P_{bq*}(\theta)}{P_{q*}(\theta,\pi_a)} \tag{6}$$

According to equation (1), every term in the first summation of the right-hand side of (6) is identically equal to one; therefore, that summation adds to Q. The second summation is equal to zero because of stationarity (4) for the likelihood maximising parameters θ. Because of (3) and considering the symmetry between the latent classes, the condition for class a applies equally in the corresponding form to class b. Then, (7) describes the balance in a model with two latent classes and constant class membership function:

$$\sum_q \frac{P_{aq*}(\theta)}{P_{q*}(\theta,\pi_a)} = \sum_q \frac{P_{bq*}(\theta)}{P_{q*}(\theta,\pi_a)} = Q \tag{7}$$

Examples in which this balance is achieved are given in Appendix A. As discussed above, the balance is broken (i.e. the optimal model contains only one latent class) when it is optimal not to include any amount of the second latent class. A diagnostic condition for this is presented in (8) and (9) for the case of a model that includes latent class a alone:

$$\left.\frac{\partial l(\theta,\pi_a)}{\partial \pi_a}\right|_{\pi_a=1} = \sum_q \frac{P_{aq*}(\theta) - P_{bq*}(\theta)}{\pi_a\, P_{aq*}(\theta) + (1-\pi_a)\, P_{bq*}(\theta)} > 0 \;\Rightarrow\; \pi_a^* = 1 \tag{8}$$

$$\sum_q \frac{P_{bq*}(\theta)}{P_{aq*}(\theta)} < Q \implies \pi_a^* = 1 \tag{9}$$

In this case of a single latent class a, $P_{q*}(\theta,\pi) \equiv P_{aq*}(\theta)$ so that $\sum_q \frac{P_{aq*}(\theta)}{P_{q*}(\theta,\pi)} = Q$. The result of Theorem 2 shows that this equality extends to each latent class in a model where two classes co-exist; this is generalised to multiple latent classes in Theorem 4 presented in section 2.2.

Conclusions from these theorems can be helpful for practitioners and researchers to explain the lack of theoretical identifiability in their models. If only one class is identified, this is not sufficient to establish that the other behaviour is absent from the data but shows only that the single class can interpret the behaviour exhibited by the other class adequately. This arises when the gain in likelihood of including a second class does not compensate the loss in likelihood for the observations that are aligned more closely to the first class.

2.1.2 The balance of latent classes with non-constant class membership function

If the class membership function π_a is not constant but is instead some function $\pi_a(\theta)$, the condition for balance is stated in Theorem 3:

THEOREM 3: *Two latent classes coexist optimally in a discrete choice model if the vector* θ *of estimated parameters satisfies the ratio specified by (11):*

$$\sum_q \frac{\frac{\partial \pi_a(\theta)}{\partial \theta} P_{aq*}(\theta) + \frac{\partial P_{aq*}(\theta)}{\partial \theta} \pi_a(\theta)}{P_{q*}(\theta)} = \sum_q \frac{\frac{\partial \pi_a(\theta)}{\partial \theta} P_{bq*}(\theta) - \frac{\partial P_{bq*}(\theta)}{\partial \theta} \left(1 - \pi_a(\theta)\right)}{P_{q*}(\theta)} \tag{11}$$

PROOF: Equation (12) states the stationarity condition required for optimality:

$$\begin{aligned} 0 &= \frac{\partial l(\theta)}{\partial \theta} \\ &= \sum_q \frac{\frac{\partial \pi_a(\theta)}{\partial \theta} P_{aq*}(\theta) + \pi_a(\theta) \frac{\partial P_{aq*}(\theta)}{\partial \theta} - \frac{\partial \pi_a(\theta)}{\partial \theta} P_{bq*}(\theta) + \left(1 - \pi_a(\theta)\right) \frac{\partial P_{bq*}(\theta)}{\partial \theta}}{\pi_a(\theta) P_{aq*}(\theta) + \left(1 - \pi_a(\theta)\right) P_{bq*}(\theta)} \end{aligned} \tag{12}$$

Equation (11) is a direct rearrangement of (12) that expresses stationarity in terms of the balance between the latent classes.

Suppose now that the set of parameters β of the class membership function is disjoint from the set θ affecting the choices themselves. Then, Theorem 3 has the following corollary:

COROLLARY 3.1: *If the class membership function, with parameters β, is independent of the latent classes, with parameters θ, the balance required of sensitivity of class membership is given by (13):*

$$\sum_q \frac{\frac{\partial \pi_a(\beta)}{\partial \beta} P_{aq*}(\theta)}{P_{q*}(\theta, \beta)} = \sum_q \frac{\frac{\partial \pi_b(\beta)}{\partial \beta} P_{bq*}(\theta)}{P_{q*}(\theta, \beta)}. \qquad (13)$$

The analysis presented in this section identifies when it is optimal for the model to include more than one latent class. Nevertheless, the coexistence of latent classes (theoretical identifiability) does not guarantee that the model will have reasonable standard deviations (empirical identifiability); we address empirical identifiability next.

2.1.3 Class behavioural diversity for empirical identifiability

To study the empirical identifiability of multiple latent classes, we assume that the model is theoretically identifiable (i.e. the model has an interior solution). If instead the model had a boundary solution (i.e. only one class was estimated), the conclusion would be that one class outperforms any combination of the two classes in explaining population behaviour.

For a parametric model to be theoretically identifiable, the information matrix *F* given in (14) must be non-singular (Rothenberg, 1971). Moreover, for greater precision in the parameter estimates, the covariance of the estimation matrix ∑ should have values on the principal diagonal with small square roots compared to the corresponding point estimates of parameters. The covariance matrix is related to the model via the Fisher information matrix *F* by (15):

$$F = -\mathbb{E}\left(\frac{\partial^2 l(\theta)}{\partial \theta_x \, \partial \theta_y}\right) \qquad (14)$$

$$\Sigma \approx F^{-1}. \qquad (15)$$

The elements on the principal diagonal of F^{-1} provide the Cramér-Rao lower bound on the variance of estimation of the parameters θ in the corresponding elements of Σ. Thus, to obtain higher precision in the estimation, the determinant of the information matrix *F* should be large, hence requiring large values of $-\mathbb{E}\left(\frac{\partial^2 l(\theta)}{\partial\theta^2}\right)$ on its principal diagonal.

As in the analysis of the first-order condition for the two-class case, we analyse the information matrix at the point determined by π_a. First, we analyse the case where the class membership function is constant. Thus, the diagonal element of the information matrix corresponding to π_a is given by the derivative of Equation (4) with respect to π_a, and relates to the empirical identifiability of the class proportions:

$$\frac{\partial^2 l(\theta)}{\partial\pi_a^2} = -\sum_q \frac{\left(P_{aq*}(\theta) - P_{bq*}(\theta)\right)^2}{\left(\mathrm{P}_{q*}(\theta)\right)^2}. \tag{16}$$

For *F* to have a large determinant, and thus for the standard errors of the estimators to be small, the magnitude of expression (16) must be large. Because the maximum likelihood estimates are obtained when the probability P_{q*}^2 is maximum, identifiability is determined by the numerator of (16). Thus, the expression $\left(P_{aq*} - P_{bq*}\right)^2$ is an essential element in the empirical identification of latent classes. Large values of this expression are obtained when the classes exhibit disparate behaviour.

Section 2.1.1 discussed how behavioural difference is needed for theoretical identifiability. For latent classes to coexist (theoretical identifiability), different preferences between the classes according to their probabilities and hence variation around 1.0 in the ratio of their model probabilities for the chosen alternative is needed. To obtain small standard deviations relative to the point estimates (empirical identifiability), the square of the difference of the latent classes must be large. Thus, empirical identifiability is promoted more by prominent behavioural contrast on a few observations rather than frequent more minor ones. Finally, note that given the addition over the sample in (16), even for small differences, as the sample size grows, the information contained also grows so that empirical identifiability increases.

2.2 Multiple Latent Class Case

We now consider the general case in which behaviour within the population aligns with several latent classes. We start by analysing the first-order conditions to generalise the results on theoretical identifiability obtained in section 2.1. Then, the analysis of empirical identifiability is extended to multiple classes.

Extending the notation of section 2.1, let π_c be the probability that individual behaviour aligns to class $c \in C$ so that $\sum_{c \in C} \pi_c = 1$ and $\pi_c \geq 0\ \forall c \in C$. Then, the joint log-likelihood function $l(\pi, \theta)$ of the model is given by (19):

$$l(\pi, \theta) = \sum_q ln\left(\sum_{c \in C} \pi_c P_{cq*}(\theta)\right) \qquad (19)$$

By extending Theorems 1 and 2, Theorem 4 establishes a necessary condition for the coexistence of several latent classes in a model:

THEOREM 4: *Several latent classes $c \in C$ coexist optimally in a model when each of them achieves the same aggregated ratio* $\sum_q \frac{P_{cq*}}{P_{q*}} = Q$.

PROOF: The likelihood (19) is maximised subject to the sum constraint $\sum_{c \in C} \pi_c = 1$ (with Lagrange multiplier λ) and positivity constraints on the probabilities $\pi_c \geq 0\ \forall c \in C$ (with Lagrange multipliers η_c) when the Lagrangian (20) is stationary with respect to variations in π_c $\forall c \in C$:

$$\mathcal{L} = -l(\pi, \theta) - \lambda\left(1 - \sum_{c \in C} \pi_c\right) - \sum_{c \in C} \eta_c \pi_c \qquad (20)$$

Differentiating the Lagrangian $\mathcal{L}$ with respect to π_c and equating to 0 for stationarity gives the necessary condition for optimality with respect to the probability π_c:

$$\frac{\partial}{\partial \pi_c}\mathcal{L} = 0 \Leftrightarrow \sum_q \frac{P_{cq*}}{\sum_{a \in C} \pi_a P_{aq*}} = \lambda - \eta_c \quad \forall c \in C$$

$$\Rightarrow \sum_q \frac{P_{cq*}}{P_{q*}} = \lambda - \eta_c \quad \forall c \in C\,.$$

The first-order Karush-Kuhn-Tucker (KKT) conditions for the positivity constraints on π_c with multiplier η_c are: $\pi_c \geq 0$, $\pi_c \eta_c = 0$, $\eta_c \geq 0$. According to the complementarity of π_c and η_c for each latent class $c \in C$,

$$\begin{aligned} \pi_c > 0 \ \Rightarrow\ \eta_c = 0 \Rightarrow \sum_q \frac{P_{cq*}}{P_{q*}} = \lambda \\ \pi_c = 0 \ \Rightarrow\ \eta_c \geq 0 \Rightarrow \sum_q \frac{P_{cq*}}{P_{q*}} \leq \lambda\,. \end{aligned} \tag{21}$$

Applying the equation for P_{q*}, the stationarity condition for likelihood and the KKT conditions gives:

$$\begin{aligned} Q &= \sum_q \frac{\sum_{c\in C} \pi_c P_{cq*}}{\sum_{a\in C} \pi_a P_{aq*}} \\ &= \sum_{c\in C} \pi_c \sum_q \frac{P_{cq*}}{P_{q*}} \\ &= \lambda \sum_{c\in C} \pi_c - \sum_{c\in C} \pi_c \eta_c = \lambda\,. \end{aligned}$$

The sum constraint $\sum_{c\in C} \pi_c = 1$ yields the value λ for the first term in the last line, whilst the KKT complementarity conditions $\pi_c \eta_c = 0 \ \forall c \in C$ yields the value 0 for the second term.

Using this in (21), $\pi_c > 0 \ \Rightarrow\ \sum_q \frac{P_{cq*}}{P_{q*}} = Q \ \forall c \in C$.

This proves Theorem 4 and extends the conclusions of the balance requirement for theoretical identifiability in section 2.1 to multiple latent classes. Those latent classes c identified by the model have identical aggregated value Q of the ratio P_{cq*}/P_{q*}; according to the second case in (21), other latent classes have aggregated values that are no greater than Q.

Theorem 4 presents the balance condition for the optimal combination of latent classes but does not guarantee their empirical identifiability. For the class membership probabilities π to be identifiable, the information matrix F should be non-singular and, because it is real and symmetric, the Hessian matrix of the Lagrangian should be positive definite. This requires that all principal submatrices of the Hessian that correspond to the second derivatives with respect to the proportions should have positive determinants. The mixed second partial derivatives of the Lagrangian $\mathcal{L}$ are equal to those of the log-likelihood (because all the constraints are linear) and are stated in (22):

$$\frac{\partial^2}{\partial \pi_a\, \partial \pi_b} \mathcal{L} = \sum_q \frac{P_{aq*} P_{bq*}}{\left(\sum_{c\in C} \pi_c P_{cq*}\right)^2} = \sum_q \frac{P_{aq*} P_{bq*}}{P_{q*}^2}\,. \tag{22}$$

Therefore, each 2×2 submatrix of this kind has the structure shown in (23):

$$\begin{bmatrix} \sum_q \frac{P_{aq*}^2}{P_{q*}^2} & \sum_q \frac{P_{aq*}P_{bq*}}{P_{q*}^2} \\ \sum_q \frac{P_{aq*}P_{bq*}}{P_{q*}^2} & \sum_q \frac{P_{bq*}^2}{P_{q*}^2} \end{bmatrix}. \tag{23}$$

Because both elements on the principal diagonal are positive, the submatrix is positive definite if the determinant exceeds zero. Moreover, if the determinant *D* given by (24) is large, then the covariances of the estimators will be small:

$$D = \sum_{p\in Q} \frac{P_{ap*}^2}{P_{p*}^2} \sum_{q\in Q} \frac{P_{bq*}^2}{P_{q*}^2} - \left(\sum_{r\in Q} \frac{P_{ar*}P_{br*}}{P_{r*}^2} \right)^2. \tag{24}$$

Before analysing (24) to assess when *D* will be positive, we note that this analysis requires that the latent classes represent distinct behaviour. Because of this, we cannot have $P_{aq*} = P_{bq*}\ \forall q$. Therefore, for each class $c \in C$ to be present there will be some cases where it outperforms the combined model. The quadratic structure of the expression $P_{cq*}^2/P_{q*}^2 \ \ c \in C$ tends to amplify the difference when one class outperforms the combined model substantially. Provided that each of the classes outperforms the combined model on some observations, then every determinant *D* of the form (24) will be positive, so that the model is theoretically identifiable. Empirical identifiability is addressed in Theorem 5.

THEOREM 5: *If several latent classes coexist in an identifiable model, empirical identifiability improves as the covariance of the latent classes decreases.*

PROOF: To make the analysis more convenient, we introduce some notation for the moments of the ratios of probabilities $\frac{P_{cq*}}{P_{q*}}\ \ c\epsilon C$. Thus, let the first and second moments be respectively:

$$\mu_c = \mathbb{E}\left(\frac{P_{cq*}}{P_{q*}}\right), c \in C$$

$$\sigma_c^2 = Var\left(\frac{P_{cq*}}{P_{q*}}\right) \ \ c \in C \ \text{ and } \ \sigma_{ab} = Cov\left(\frac{P_{aq*}}{P_{q*}}, \frac{P_{bq*}}{P_{q*}}\right) \quad a, b \in C\ .$$

With this notation, the expectation of elements in (24) can be written as:

$$\mathbb{E}\left(\sum_{q\in Q}\frac{P_{cq*}^2}{P_{q*}^2}\right)=Q(\mu_c^2+\sigma_c^2)\ \text{ and }\ \mathbb{E}\left(\sum_{q\in Q}\frac{P_{aq*}P_{bq*}}{P_{q*}^2}\right)=Q(\mu_a\mu_b+\sigma_{ab})\,.$$

Therefore, the expectation of (24) can be rearranged to express *D* as an unbiased sample estimate of the population quantity:

$$\frac{1}{Q^2}\mathbb{E}(D)=\mu_a^2\mu_b^2\left(\frac{\sigma_a^2}{\mu_a^2}-2\frac{\sigma_{ab}}{\mu_a\mu_b}+\frac{\sigma_b^2}{\mu_b^2}\right)+\sigma_a^2\sigma_b^2\left(1-\frac{\sigma_{ab}^2}{\sigma_a^2\sigma_b^2}\right) \tag{25}$$

Recall that from condition (21), for both classes a and b to be present in the model we need $\mu_a=\mu_b=1$. If the choice probabilities are perfectly correlated, then $\sigma_{ab}^2=\sigma_a^2\sigma_b^2$ so that the second term on the right-hand side of (25) would be null. The remaining term would then be $(\sigma_a-\sigma_a)^2$ with perfect correlation and neither class dominating the other, this will also be null. The Hessian matrix would therefore be singular in expectation. The expectation of the partial derivative of *D* with respect to the correlation σ_{ab} in (26) is negative so that the expectation of the determinant increases as this correlation decreases. In particular,

$$\mathbb{E}\left(\frac{\partial D}{\partial\sigma_{ab}}\right)=-2Q^2(\mu_a\mu_b+\sigma_{ab})=-2Q^2\mathbb{E}\left(\frac{P_{aq*}P_{bq*}}{P_{q*}^2}\right)\leq 0\,. \tag{26}$$

Consequently, estimation of the mixed model is better conditioned (as indicated by larger *D* values) when correlation σ_{ab} decreases and as the sample size *Q* increases, thus proving Theorem 5.

Therefore, the requirement for positive determinants of the principal submatrices of the Hessian generalises the requirement for the binary classes' case presented in section 2.1. To be identifiable, the behaviour of a class should outperform that of the combined model in at least one observation; the greater the behavioural difference, the greater the determinant (24) and hence the smaller the covariance of the estimators.

In conclusion, Theorem 4 presents the balance conditions required if the presence of several latent classes is optimal. Theorem 5 generalises the requirements for empirical identifiability showing in a simple structure that empirical identifiability increases as the behavioural difference of the latent classes increases as quantified by decreasing covariance among them.

3. EXEMPLIFYING FACTORS AFFECTING EMPIRICAL EXPERIMENTS IN A REALISTIC CONTEXT

We now show how the factors identified by the theorems developed in Section 2 apply in practice. Specifically, we focus on outcomes and conclusions from Sections 2.1.1 and 2.1.3, where we addressed the conditions for theoretical and empirical identifiability respectively in the two-class context. Our objective is to show how different drivers of identifiability are related to the theoretical background that we have established. Conclusions from these experiments can help practitioners understand potential causes of non-identifiability, how this relates to the over-arching theorems, and to understand the implications of this for survey design and analysis.

We chose as testing ground the case of multiple-choice heuristics. As presented in Section 1, this context usually provides challenging identifiability scenarios (e.g. Leong and Hensher, 2012; Hess and Stathopoulos, 2013). Here, each choice heuristic is modelled under a different latent class.

In the experiment formulation, to guarantee the presence of different choice heuristics and control the choice parameters, we generated a synthetic population following the seminal work of Williams and Ortúzar (1982). We tested three dimensions affecting the choice process that could potentially affect identifiability: (i) the latent class behaviour given by a distinct choice heuristic, which will determine the behavioural difference quantified in equation (16) and hence empirical identifiability; (ii) the proportion of each latent class in the synthetic sample, which will affect the feasibility of equation (7), determining the existence of balance; and (iii) the correlation between the parameters of the probability of belonging to each class and the parameters associated with their sensitivities for different attributes of the alternatives, which could provide external confounding effects exemplifying a more general case. For each case of these three dimensions, ten simulation experiments were performed.

The first dimension described is the latent class formulation, in our case, given by the choice heuristic. The analysis of Section 2 established that the difference between the latent classes is key to their identification as quantified in equation (16). Three different choice heuristics were tested against random utility maximisation (RUM), the most widely used, to investigate whether they could be identified in our practical context. These are: Elimination by Aspects –EBA– (Tversky, 1972a; 1972b), Stochastic Satisficing –SS– (González-Valdés and Ortúzar, 2018) and Random Regret Minimisation –RRM– (Chorus et al., 2008).

The second dimension is the proportion of each latent class (or choice heuristic) in the sample. The results (5) and (7) show that the greater this proportion, the greater the number of observations for which one latent class will outperform the other, thus increasing its presence in the balance. Two proportions were tested: 70% of the sample chooses according to RUM and 30% according to the other heuristic, and *vice versa*, that is, $\pi_c \in \{0.3, 0.7\}$.

Finally, the third dimension is the correlation between the choice and the class membership probabilities. This dimension aims to analyse how any such correlation would affect identifiability. This correlation was introduced through a *personal trait* that affects both the probability of belonging to a class and the choice preferences.

We use a simulated dataset to investigate how these factors affect the theoretical and empirical identifiability in a realistic context. For estimation, we require two components: for each individual a set of alternatives available and their choices from this set. The choice sets for the individuals were extracted from a revealed preference dataset to represent a realistic scenario; the individuals' choices were simulated for the synthetic population under the various heuristics to control the underlying behaviours.

3.1 The Choice Sets

The choice sets were created based on a well-tested dataset from a transport survey in Santiago de Chile (Gaudry et al., 1989; Guevara, 2016; Jara-Díaz & Ortúzar, 1989), comprising the trips from home to work of 1,374 individuals, who chose among a maximum of nine modes.

This dataset provided real choice sets ranging from two to nine alternatives from which the simulated choice sets were created. To control the number of alternatives available in the experiment, all choice sets presented to our synthetic individuals were specified with three alternatives. Moreover, we could also estimate the respective alternative-specific constants (ASC) because the alternatives were labelled.

Two separate processes were performed to create the simulated choices: (i) fictitious choice sets of size 3 were generated and (ii) each individual's choice was simulated for each one of these sets. To generate these fictitious choice sets, real choice sets were sampled from the databank and then adjusted as follows. If the sampled choice set had fewer than three alternatives, it was discarded;

if it had more than three alternatives, one of the alternatives was deleted at random[4]. This process was repeated[5] until the choice set size was reduced to three.

After the choice sets were generated, each individual's choice was simulated under the specified heuristic. Each alternative in the choice sets was characterised by four attributes: monetary cost, in-vehicle time, walking time, and waiting time.

3.2 Synthetic Population and Choice Heuristics

We followed four steps to simulate the choice of an alternative from the simulated choice sets. First, we created the individuals' traits. To do this, a binary variable was generated for each individual in the sample to represent their socio-demographic attribute z (named *trait*) with probability p_z. Second, each simulated individual was assigned independently to use one of the two available choice heuristics: RUM and the contrasting one (i.e. EBA, RRM or SS). These choice heuristics are explained in more detail below. In each case, the probability π_R of using RUM was given by the inverse logit function (27) with parameters shown in Table 1.

$$\pi_R = \frac{\exp(\theta_0 + \theta_1 z)}{1 + \exp(\theta_0 + \theta_1 z)} \tag{27}$$

Following this, a choice set was selected from the simulated databank of 28,477. Finally, the individual's choice from their choice set was simulated according to their assigned heuristic.

[4] This elimination process is agnostic to which alternative was chosen in the real context. Thus, although the real chosen alternative might be eliminated, this does not create any difficulty because we simulated from the choice sets that we generated.

[5] Because we delete excess alternatives at random to generate each choice set, an initial set of size 4, say, can create four different choice sets of size 3 (by deleting a different alternative in each one); whereas one of size 9 can create 84 different choice sets of size 3. Accounting for all the sets in the original dataset, we had a total of 28,477 different choice sets to pool from. We repeated this procedure of uniform random sampling with replacement from the 1,374 individuals to generate a synthetic sample of 10,000 choices.

Table 1. Synthetic population latent class parameters

Parameter	Value
θ_0[6]	0
θ_1	+/- 1.39
p_z	0.70

Random utility maximisation (RUM)

RUM is the most widely used heuristic in choice modelling. We used its simplest form, the multinomial logit model – MNL – (McFadden, 1973) with additive linear in the parameters utility function. In some experiments, the cost attribute was modified based on the individual's sociodemographic *trait* to test the effect of correlation between the class membership function and the choice heuristic. If the individual had the trait (indicated by $z = 1$), the sensitivity to cost was modified; we called this attribute *cost difference* of sensitivity. The model parameters used for this simulation are given in Appendix B.

Random regret minimisation (RRM)

RRM (Chorus et al., 2008) is a heuristic where individuals evaluate alternatives relative to each other. It is based on the concept of *anticipated regret*, which is the feeling stimulated when the individual imagines what they would have experienced if they had chosen another alternative (Simonson, 1992). Among the several versions of RRM, we considered the $\mu - RRM$ (van Cranenburgh et al., 2015), where the regret R_i for each alternative i is given by (28):

$$R_i = \sum_{j \in J,\, j \neq i} \sum_{k \in K} \mu \log_e \left(1 + \exp\left(\frac{\beta_k}{\mu} \left(x_{jk} - x_{ik} \right) \right) \right) \tag{28}$$

The parameter μ in this formulation controls the *profundity of regret*: smaller values represent emphasised regret and strengthened preference for the most attractive alternative. We selected this formulation to increase the profundity of regret compared to the simplest version, which implicitly

[6] Note that in our case θ_0 is not necessary for simulation. However, a modeller unaware of the function that was used to generate the probabilities (which would normally be the case) could test a model considering it. If the model is estimated correctly, this parameter will not differ significantly from zero.

has μ=1 (Chorus, 2010). The greater profundity of regret induced by using the value μ=0.2 increases the behavioural difference between RRM and RUM. According to Theorem 5, increasing the difference between the choice heuristics increases the chance of identifying them jointly. The model parameters used for the simulation are also given in Appendix B.

Stochastic Satisficing (SS)

Satisficing is a bounded rationality heuristic that involves several simplifications to rational decision-making (Simon, 1955; 1956). Because Simon's definition is incompletely detailed, several interpretations of this theory exist, and no consensus has yet been reached about the precise definition (Manski, 2017). Here, we interpret satisficing as a heuristic according to which individuals choose the first satisfactory (i.e. good enough) alternative they consider.

Among several possible implementations of this heuristic, we use the *Stochastic Satisficing –SS–* model (González-Valdés and Ortúzar, 2018), where the probability (29) of an alternative *i* being acceptable for an individual q is the product of the probabilities that each of the attributes *k* is acceptable (30):

$$Pr\left(A_{iq}=1\right)=\prod_{k} a_{kiq} \tag{29}$$

where

$$a_{kiq}=\frac{\exp\left(\beta_k\left(x_{kiq}-f_k\right)\right)}{1+\exp\left(\beta_k\left(x_{kiq}-f_k\right)\right)}. \tag{30}$$

The probability a_{kiq} of each attribute *k* of alternative i being acceptable to individual q is given by the logistic function (30), where β_k represents the sensitivity to attribute *k* and f_k is the associated acceptability reference (threshold) value for that attribute. In this model, different attributes may appear in the acceptability functions of the various alternatives. Even though theoretically, sensitivities and thresholds functions can vary across alternatives and individuals, we modelled the simplest version with constant sensitivities and thresholds among alternatives.

In our simulation, costs were modelled by separate acceptability functions, whilst in-vehicle, waiting, and walking time were modelled using the same acceptability function in each alternative. For the time attribute, a time sensitivity, a time reference value, and two marginal rates of substitution were estimated. The marginal rates of substitution represent the equivalence between travel time and, respectively, waiting time or walking time. The values used for all parameters in the simulation are specified in Appendix B. To simulate choices under this interpretation of stochastic satisficing, the alternatives were sampled with replacement from the choice set and evaluated for acceptability according to (29) until one was accepted.

Elimination by aspects (EBA)

EBA (Tversky, 1972a, 1972b) is a bounded rationality choice heuristic where individuals consider alternatives according to the values of a sequence of attributes, following a recursive procedure[7]. At each step, individuals select the most important (to them) of the remaining aspects (attributes) and discard every alternative that does not satisfy pre-specified thresholds. This process continues until only one option remains, which is therefore selected.

Each aspect ($k \in K$) has an associated weight (w_k) which determines the probability of being considered in the decision process. The modelling process adopted here estimates the logarithm α_k of each weight w_k (32) in the whole real space.

$$P_k = \frac{w_k}{\sum_j w_j} \tag{31}$$

$$w_k = \exp(\alpha_k) \tag{32}$$

The weights used for each aspect are given in Appendix B.

In the EBA model, selection according to each aspect is binary. Although the attributes may have continuous values as in the present case, the acceptability thresholds are specified to achieve binary discrimination. We considered two thresholds for the cost attribute, at US$ 0.25 and US$ 0.65 (i.e. three aspect levels were created with two of them considered desirable). Whereas for travel time, waiting time and walking time, one threshold[8] for each attribute was adopted at 15, 5 and 3 mins,

[7] Examples of simple EBA models can be found in the work of Gilbride & Allenby (2006).

[8] Note that this is indeed a "threshold" that corresponds to a critical value that determines acceptability. By contrast, the reference value in SS has a continuous influence on the probability of acceptability.

respectively. Therefore, an alternative would be discarded if any one of its time elements exceeded the corresponding threshold.

3.3 Estimation Procedure

Previewing the results detailed in Section 4, we found different degrees of identifiability in our models, which is a consequence of the borderline cases we designed in our experiment. In some cases, models were strictly identifiable, meaning all parameters were recovered (following the definition proposed by Gu and Xu, 2020). In contrast, models were only partially identifiable in some other cases, meaning that a combination of model parameters allowed for observationally equivalent models. Finally, in the remaining cases, estimation resulted in a single component without a balance being achieved.

For the specific context of testing the theorems on borderline identifiable cases, we preferred Bayesian estimation over maximum likelihood estimation for several reasons. First, fundamental non-identifiability in maximum likelihood is related to the impossibility of inverting the Hessian matrix of the likelihood, with identifiability being a diagnosis of the estimation. Misdiagnosis of non-identifiability can be related to any one or more of (i) failure of the maximisation algorithm, (ii) the algorithm used to calculate the information matrix, (iii) assumption of asymptotical convergence to the covariance matrix or indeed, (iv) a fundamental lack of identifiability.

On the other hand, in Bayesian estimation procedures, non-identifiability is inferred from the estimators. We assessed identifiability based on the ratio of the standard deviation of the posterior to the prior, which is a measure of progress in the Bayesian estimation. We also considered the variation among simulations of the estimates of parameter values, which quantifies the consistency in the outcomes of the Bayesian estimation process. Misdiagnosis of false non-identifiability[9] under Bayesian estimation can result if the priors are too narrow or the Markov chains used for estimation are too short to cover the whole posterior distribution effectively. We protected against misdiagnosis by using wide-uninformative priors and allowing long Markov chains to run in the

[9] Bayesian estimation is prone to the opposite, misdiagnosis of false identifiability due to inadequate priors. Narrow or over-informative priors could identify the model by erroneously eliminating potential plausible parameter combinations.

testing scenarios[10]. Notwithstanding this, we note that Train (2009, p290) showed that the choice of estimation technique as between likelihood maximisation and Bayesian estimation has little influence when the sample size Q is large (as in our case).

Previous tests have suggested that this estimation procedure usually requires numerous iterations to achieve stationarity (Godoy and Ortúzar, 2008). Here, 5,000 burn-in samples were discarded before sampling from the Markov chain. After this, 10,000 samples were obtained from the posterior distribution of the parameters.

4. ANALYSIS OF RESULTS

Given the combinations of dimensions tested and the replications for each combination, a total of 120 experiments were undertaken. In each of these experiments, we simulated the choices of a sample of 10,000 individuals and then estimated the choice models from the resulting data. First, across the various dimensions, we analysed the proportion of replications of each model that resulted in a balance between the latent classes. Then, we verified that Theorem 2 held for the models that identified both latent classes notwithstanding being estimated using Bayesian methods[11].

4.1 Analysis of Convergence

In this section, we report the progress of the Bayesian estimation process by analysing the posterior distribution. In practice, the target distribution of the parameters will not usually be known; this has led to the formulation of measures such as the potential scale reduction factor (PSRF), which compares the variability of parameter estimates between Markov chains with that within them (Gelman and Rubin, 1992; Brooks and Gelman, 1998). Values of the resulting test statistic close to 1.0 indicate chain convergence, with variability between chains consistent with that within them. PSRF values greater than 1.0 show excessive variability between chains, which indicates lack of convergence due either to insufficient sampling or non-identifiability. We calculated this by

[10] Bayesian estimation was undertaken using Markov Chain Monte Carlo, specifically Gibbs sampling, using the JAGS package (Plummer, 2016) for the R software system (R Core Team, 2016).

[11] For reference and comparison, we also report results of maximum likelihood estimation of two models on identical simulated datasets in Appendix D.

comparing the long-term variability of parameter estimates with their short-term variability[12]. In the present context, however, parameter estimates could converge to those of a single latent class so that a satisfactory PSRF value is not sufficient for model recovery.

The mean over 10 simulations of the PSRF is shown in Table 2 together with the standard deviation for each of the 12 scenarios tested. This was considered acceptable in 8 of the 12 scenarios, with values in the range [1, 1.08] and standard deviations no greater than 0.08. The remaining four scenarios had mean PSRF values of 1.16 or greater, with standard deviations ranging from 0.39 to 1.21. In these four scenarios, the mean values exceeding the reference value of 1.10 show instability in some of the estimation processes, whilst the large standard deviations show further that the instability varied among the 10 simulations. Together, these indicate lack of reliable convergence in the estimation of the four models with those combinations of choice heuristics. We explore the implications of this for model identifiability in the next section.

Table 2. Potential scale reduction factor

	Case Mean (Standard Deviation); Median across ten cases per scenario			
	71% RUM class		71% Non-RUM class	
	Class correlation		Class correlation	
Latent classes	None	Positive	None	Positive
RUM & RRM	1.08 (0.08); 1.05	1.16 (0.39); 1.04	1.65 (0.82); 1.37	1.51 (1.21); 1.11
RUM & SS	1.07 (0.05); 1.05	1.39 (0.91); 1.06	1.05 (0.03); 1.05	1.06 (0.04); 1.06
RUM & EBA	1.06 (0.04); 1.05	1.08 (0.08); 1.06	1.06 (0.03); 1.06	1.07 (0.05); 1.05

4.2 Analysis of Identifiability

A model is non-identifiable if the information matrix is singular, which is equivalent to having an infinite element within the covariance matrix. In our context of Bayesian estimation, no matrix inversion is required; nevertheless, model non-identifiability can be detected when the standard

[12] Because we ran only one chain for each of the estimations, we estimate the scale reduction factor by cutting each post-burn-in chain into two subsets of 5,000 samples each.

deviations of the posterior distributions of the parameters are extreme with associated instability of the Markov chain, which is illustrated in the example in Appendix C and consistent with section 4.1. This instability is manifest in failure of the posterior distribution of parameters to develop from the initial prior, leading to excessive variability between estimates of different Markov chains relative to that within the chains. In our investigation, we adopted accuracy of parameter recovery as an indicator of model identifiability.

Following Gu and Xu (2020) to distinguish degrees of identifiability that models may exhibit, we developed three further descriptions:

- *Strict identifiability*: all model parameters are estimated with acceptable standard deviations. Both latent classes are identified, thus there is a balance between them.
- *Partial identifiability*: the model balances two classes (theoretical identifiability) and model parameters are estimated accurately, but a small proportion of them are estimated with extreme standard deviations (empirical non-identifiability).
- *Non-identifiability*: most parameters are estimated with extreme standard deviation, or no balance can be found between latent classes.

In section 2, we analysed how behavioural differences may impact the identifiability of the latent classes. Figure 1 provides a graphical diagnostic that shows the distribution of behavioural differences between the RUM class and the other choice heuristic classes among the alternatives of the dataset. This is quantified by the absolute difference between the probabilities given by the two choice heuristics. For example, if two heuristics a and b estimate probabilities P_{ai} and P_{bi} of choosing alternative i, then the difference is calculated as $|P_{ai} - P_{bi}|$.

Figure 1 shows that among the cases that we considered, the RRM latent class differs least from the RUM latent class in its behaviour. Thus, we expect the RRM latent class to have the least chance of achieving balance with RUM in this context. Conversely, each of SS and EBA represents a substantial behavioural difference from the RUM latent class. Note, however, that because this analyses only one dimension of the information matrix, it helps generate hypotheses but does not guarantee universal support for them.

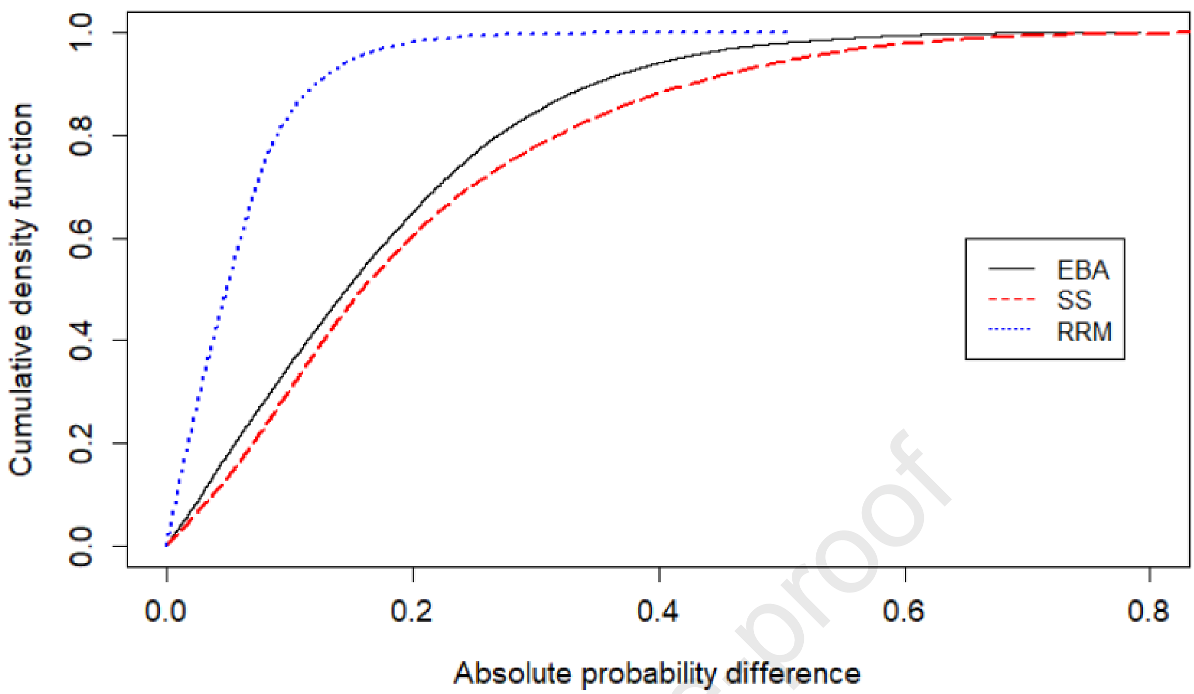


Figure 1 Behavioural difference between RUM and each of RRM, SS and EBA

We analysed each pair of latent classes separately and evaluated the results according to the three degrees of identifiability. We also analysed separately the influence of the population proportions in each of the latent classes simulated and each correlation case as follows.

RRM and RUM classes identifiability analysis

Table 3 shows the results of the identifiability analysis for each of the 40 estimations among the four scenarios of correlation and proportions of each latent class in the combination. In most cases in which the sample was dominated by the RUM latent class, it was the only one identified and no balance was achieved between the RUM and RRM classes in any of the cases. Linking to the theorems provided, this suggests that even though some individuals in the simulated population exhibit RRM behaviour, the improvement in model fit by including an additional class for them is insufficient to compensate for the consequent worsening of fit for the RUM individuals.

When the RRM class dominated the sample, its identifiability increased, although it was less identifiable than the RUM class when that dominated. When no correlation was present between the class membership function and the parameters of the RUM class, RRM was identified in seven of the ten cases. Nevertheless, in three of the seven cases where the RRM class was identified, it was identified weakly, with some parameters having extreme variance.

Table 3. Identifiability results of RUM vs RRM models

Correlation	RUM dominates π_R = 0.71	RRM dominates π_R = 0.29
No correlation	$\frac{8}{10}$ identifies RUM only $\frac{2}{10}$ identifies RRM only **No balance detected**	$\frac{3}{10}$ identifies RUM only $\frac{4}{10}$ identifies RRM only $\frac{3}{10}$ identifies partially RRM **No balance detected**
Positive correlation	$\frac{9}{10}$ identifies RUM only $\frac{1}{10}$ identifies RRM only **No balance detected**	$\frac{3}{10}$ identifies RUM only $\frac{6}{10}$ identifies RRM only $\frac{1}{10}$ identifies RRM partially **No balance detected**

Finally, when the correlation between the class membership function and the RUM heuristic was greater, the strength of identifiability of the RRM increased with only one case still being partially identifiable; this can be understood as being due to the increased difficulty in identifying the RUM class.

The PSRF values calculated for this combination of latent classes are shown in the first row of Table 2. Although the PSRF value of 1.08±0.08 indicates a good degree of convergence in the un-correlated case where RRM dominates, the results in Table 3 show that this arose because of convergence of a model form with a single latent class. The remaining three cases of mixed RUM and RRM classes all had high PSRF values, ranging from 1.16±0.39 to 1.65±0.82 indicating lack of convergence of the estimation. In none of these four cases was a balanced combination of latent classes identified.

The results from these cases, whichever RUM or RRM dominates, are consistent: the balance, or coexistence of RUM and RRM in the estimated choice model, is improbable in this dataset. Based on inference from the relationships (8), this suggests that in our simulated mode choice dataset, the RUM mechanism seems to be more robust in that it can accommodate RRM individuals better than can the RRM accommodate RUM ones, thus emerging from the estimation more frequently than RRM. Therefore, under no balance conditions, identifiability is not expected.

Moreover, note that although we used the μ-RRM to increase the probability of detecting coexistence by emphasising the behavioural difference between RRM and RUM, this was not sufficient for effective identifiability. We also tested an increased sample size; but even a sample size of 40,000 observations did not provide enough information to identify these two classes in balance. However, several authors have reported identifying RRM and RUM jointly in practice without the need for latent variables (e.g. Boeri et al, 2014; Boeri and Longo, 2017). Thus, we conclude that the lack of behavioural difference exhibited in the choice scenarios presented here is a good indication of the plausibility of identification (consistent with Figure 1).

SS and RUM identifiability analysis

Table 4 shows the results of estimating SS and RUM jointly. In the cases where the RUM class dominated the sample, it was always identified with a degree of balance in the model. The SS class was identified only weakly, because some parameters had extreme variance. The mean PSRF of 1.07 for this case in Table 2 shows that the estimates converged. Introducing greater correlation did not affect identifiability, though it did reduce the convergence of estimation as quantified by the large mean PSRF of 1.39.

Table 4. Identifiability results of RUM and SS models

Correlation	RUM dominates π_R = 0.71	SS dominates π_R = 0.29
No correlation	$\frac{9}{10}$ identifies RUM and weakly SS $\frac{1}{10}$ weakly identifies RUM and SS **Partial identifiability and balance detected**	$\frac{10}{10}$ identifies RUM and SS **Balance and identifiability detected**
Positive correlation	$\frac{9}{10}$ identifies RUM and weakly SS $\frac{1}{10}$ identifies weakly RUM only	$\frac{9}{10}$ identifies RUM and SS $\frac{1}{10}$ identifies RUM and weakly SS

	Partial identifiability and balance detected	Balance and identifiability detected

When the SS class dominated the sample, a proper balance was detected whilst the mean PSRF of 1.06 – 1.08 shows good convergence of estimation. The model was able to identify the estimators of the RUM class, the SS class, and the class membership function with reasonably small variance. When correlation was introduced, the degree of identifiability decreased slightly.

These results for the SS and RUM latent classes show that a balance can be achieved in this model, although it depends on the proportion of the population that uses each of these choice heuristics. When most individuals followed the RUM class, incorporating the SS class did not usually compensate for the loss of likelihood of the RUM individuals. Conversely, when the proportion of SS dominated, the better performance of the RUM individuals did compensate for the decrease in the likelihood for the SS individuals. Hence, a balance may be achieved when SS individuals are more numerous than RUM ones. In either case, the RUM appears to be the more robust heuristic in our context, because it could be identified even in cases where it was present in low proportion.

As in the case of RRM, we also tested increasing the sample size to 40,000 observations. In all these cases, we detected strong identifiability of both latent classes. Thus, the increase in information was sufficient for identifiability.

We note, again, that these results are specific to the present dataset. However, if a dataset provides choice situations in which SS behaviour differs more from RUM – and the individuals behave following such heuristics – achieving balance seems possible.

EBA and RUM identifiability analysis

Table 5 shows the results of estimating the Latent Class model with RUM and EBA as choice heuristics. These results show that balance was achieved in the 40 experiments, with mean PSRF values in the range 1.05 – 1.06 (Table 2), showing good convergence. In all cases the dominant latent class was identified accurately, as was the other class in most cases. In a few cases, the minority class was identified only weakly, with slightly more of these when there was positive correlation.

These results show that in our scenarios, when RUM and EBA are present in the data, they can be identified jointly, indicating that neither RUM nor EBA can represent the behaviour of the other choice heuristic effectively.

Table 4. Identifiability results of RUM and EBA models

Correlation	RUM dominates $\pi_R = 0.71$	EBA dominates $\pi_R = 0.29$
No correlation	$\frac{9}{10}$ identifies RUM and EBA $\frac{1}{10}$ identifies RUM and weakly EBA **Balance and identifiability detected**	$\frac{10}{10}$ identifies RUM and EBA **Balance and identifiability detected**
Positive correlation	$\frac{7}{10}$ identifies RUM and EBA $\frac{3}{10}$ identifies RUM and weakly EBA **Balance and identifiability detected**	$\frac{9}{10}$ identifies RUM and EBA $\frac{1}{10}$ identifies EBA and weakly RUM **Balance and identifiability detected**

5. CONCLUSIONS

Latent class (LC) models have been reported in the literature for several decades. In these models, identifiability is key to determining whether the different classes are present in the data. For this kind of model, identifiability has only been studied in general terms, whilst empirical identifiability has not been considered in depth. This paper presents theoretical and empirical studies of identifiability for LC models.

The theoretical framework developed here of LC models provides a basis for analysis of their identifiability. Through this, we established two analytical conditions for identifiability. First, there must be a balance between the latent classes for theoretical identifiability. Second, the behaviour of the classes must differ sufficiently so that they can be identified empirically with acceptable accuracy in estimates of their parameters. The balance required for joint estimation requires that one latent class is not sufficiently good in explaining the behaviour of members of the other class, which we quantify in the balance equation. On empirical identifiability, we show that the latent classes must differ sufficiently in their typical behaviour and that the data used in estimation must include sufficient cases that expose this difference. If either of these conditions is not satisfied, then simultaneous identification of the latent classes in a single model will not be possible.

To show how the theoretical framework developed here links to practical scenarios, we tested it using data synthesised for choice situations using three pairs of choice heuristics: Random Utility Maximisation (RUM) in combination with each of Elimination by Aspects (EBA), Random Regret Minimisation (RRM) and Stochastic Satisficing (SS). Each of these three combinations of choice models were estimated using Bayesian statistical methods. For each mixture, 40 cases were simulated in four groups of 10 that differed in correlation and choice heuristic dominance.

Our experiments show that estimation may fail to identify both classes, even though the generating process contains a mixture of them. The existence of a balance depends on the inadequacy of each class in representing the behaviour of the other in some cases. Indeed, the dominant heuristic must perform poorly in some cases following the other to be able to estimate the model fully.

In view of these findings, a practical strategy would be to analyse the classes before estimating a combined model. This can be undertaken using the straightforward diagnostic tests presented here. This way, using some testing parameters, modellers can examine whether the datasets are sufficiently rich in their choice behaviour to support joint estimation of the desired heuristics.

Finally, we note that the empirical analysis was made for a specific context and limited number of alternatives. A latent class that could not be identified in a particular context might be suitable and identifiable in another context that provides sufficient richness to expose the behavioural differences.

ACKNOWLEDGEMENTS

We are grateful to the Chilean National Commission of Scientific and Technological Research (CONICYT 21151280) for supporting this research. We are also grateful to the Instituto Sistemas Complejos de Ingeniería (CONICYT PIA/BASAL: AFB18003), the BRT+ Centre of Excellence (funded by the Volvo Research and Educational Foundations) and the Centre for Sustainable Urban Development, CEDEUS (CONICYT/FONDAP/15110020). The authors acknowledge the use of the UCL Myriad High Throughput Computing Facility (Myriad@UCL) and associated support services in the completion of this work.

REFERENCES

Adamowicz, W.L., Swait, J.D., 2013. Are food choices really habitual? Integrating habits, variety-seeking, and compensatory choice in a utility-maximizing framework. American Journal of Agricultural Economics 95, 17–41.

Araña, J.E., León, C.J., Hanemann, M.W., 2008. Emotions and decision rules in discrete choice experiments for valuing health care programmes for the elderly. Journal of Health Economics 27, 753–769.

Balbontin, C., Hensher, D.A., Collins, A.T., 2017. Integrating attribute non-attendance and value learning with risk attitudes and perceptual conditioning. Transportation Research Part E: Logistics and Transportation Review 97, 172–191.

Ben-Akiva, M., Boccara, B., 1995. Discrete choice models with latent choice sets. International Journal of Research in Marketing 12, 9–24.

Boeri, M., Longo, A., 2017. The importance of regret minimization in the choice for renewable energy programmes: evidence from a discrete choice experiment. Energy Economics 63, 253-260.

Boeri, M., Scarpa, R., Chorus, C.G., 2014. Stated choices and benefit estimates in the context of traffic calming schemes: utility maximization, regret minimization, or both? Transportation Research part A: Policy and Practice 61, 121-135.

Brooks, S.P., Gelman, A., 1999. General methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics 7, 434-55.

Cherchi, E. and Ortúzar, J. de D., 2008. Empirical identification in the mixed logit model: analysing the effect of data richness. Networks and Spatial Economics 8, 109-124.

Chiou, L., Walker, J.L., 2007. Masking identification of discrete choice models under simulation methods. Journal of Econometrics 141, 683–703.

Chorus, C.G., 2010. A new model of random regret minimization. European Journal of Transport and Infrastructure Research 10, 181–196.

Chorus, C.G., Arentze, T.A., Timmermans, H.J.P., 2008. A random regret-minimization model of travel choice. Transportation Research Part B: Methodological 42, 1–18.

Gaudry, M.J.I., Jara-Diaz, S.R., Ortúzar, J. de D., 1989. Value of time sensitivity to model specification. Transportation Research Part B: Methodological 23, 151–158.

Gelman, A., Rubin, D. 1992. Inference from iterative simulation using multiple sequences. Statistical Science 7, 457-511.

Gilbride, T.J., Allenby, G.M., 2006. Estimating heterogeneous EBA and economic screening rule choice models. Marketing Science 25, 494–509.

Godoy, G., Ortúzar, J. de D., 2008. On the estimation of mixed logit models. In: Inweldi, P.O. (Ed.), Transportation Research Trends. Nova Science Publishers, New York, pp 289-310.

González-Valdés, F., Ortúzar, J. de D., 2018. The stochastic satisficing model: a bounded rationality discrete choice model. Journal of Choice Modelling 27, 74–87.

Gonzalez-Valdes, F., Raveau, S., 2018. Identifying the presence of heterogeneous discrete choice heuristics at an individual level. Journal of Choice Modelling 28, 28–40.

Greene, W.H., Hensher, D.A., 2003. A latent class model for discrete choice analysis: contrasts with mixed logit. Transportation Research Part B: Methodological 37, 681-698.

Gu, Y., Xu, G. 2020. Partial identifiability of restricted latent class models. The Annals of Statistics 48, 2082-2107.

Guevara, C.A., 2016. Mode-valued differences of in-vehicle travel time savings. Transportation 44, 977–997.

Guevara, C.A., Chorus, C.G., Ben-akiva, M.E., 2016. Sampling of alternatives in random regret minimization models. Transportation Science 50, 306–321.

Hess, S. Chorus, C.G. 2015. Utility maximisation and regret minimisation: a mixture of a generalisation. In: Rasouli, S. Timmermans, H. (Eds.), Bounded Rational Choice Behaviour: Applications in Transport. Emerald, Bingley, pp. 31-47.

Hess, S., Stathopoulos, A., 2013. A mixed random utility - random regret model linking the choice of decision rule to latent character traits. Journal of Choice Modelling 9, 27–38.

Hess, S., Stathopoulos, A., Daly, A., 2012. Allowing for heterogeneous decision rules in discrete choice models: an approach and four case studies. Transportation 39, 565–591.

Hsiao, C., 1983. Identification. In: Griliches, Z., Intrilligator, M. (Eds.), Handbook of

Econometrics. North Holland, Amsterdam, pp. 223–283.

Huang, G.-H., Bandeen-Roche, K., 2004. Building an identifiable latent class model with covariate effects on underlying and measured variables. Psychometrika 69, 5–32.

Jara-Díaz, S.R., Ortúzar, J. de D., 1989. Introducing the expenditure rate in the estimation of mode choice models. Journal of Transport Economics and Policy 23, 293–308.

Kamakura, W.A., Russell, G.J., 1989. A probabilistic choice model for market segmentation and elasticity structure. Journal of Marketing Research 26, 379-390.

Leong, W., Hensher, D.A., 2012. Embedding multiple heuristics into choice models: an exploratory analysis. Journal of Choice Modelling 5, 131–144.

Manski, C.F., 2017. Optimize, satisfice, or choose without deliberation? A simple minimax-regret assessment. Theory and Decision 83, 155–173.

Matzkin, R.L., 2007. Nonparametric identification. In: Heckman, J., Leamer, E. (Eds.), Handbook of Econometrics. North Holland, Amsterdam, pp. 5307–5368.

McElreath, R., 2015. Statistical Rethinking: A Bayesian Course with Examples in R and Stan. CRC Press, Boca Raton.

McFadden, D., 1973. Conditional logit analysis of qualitative choice behavior. In: Zarembka, P. (Ed.), Frontiers of Econometrics. Academic Press, New York, pp. 105-142.

McNair, B.J., Hensher, D.A., Bennett, J., 2012. Modelling heterogeneity in response behaviour towards a sequence of discrete choice questions: a probabilistic decision process model. Environmental and Resource Economics 51, 599–616.

Ortúzar, J. de D., Fernández, J.E., 1985. On the stability of discrete choice models in different environments. Transportation Planning and Technology 10, 209–218.

Plummer, M., 2016. RJags: Bayesian Graphical Models using MCMC. R package version 4-6.

R Core Team, 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rossetti, T., Guevara, C.A., Galilea, P., Hurtubia, R., 2018. Modeling safety as a perceptual latent variable to assess cycling infrastructure. Transportation Research Part A: Policy and Practice

111, 252–265.

Rothenberg, T., 1971. Identification in parametric models. Econometrica 39, 577–591.

Simon, H.A., 1955. A behavioral model of rational choice. The Quarterly Journal of Economics 69, 99–118.

Simon, H.A., 1956. Rational choice and the structure of the environment. Psychological Review 63, 129–138.

Simonson, I., 1992. The influence of anticipating regret and responsibility on purchase decisions. Journal of Consumer Research 19, 105–118.

Swait, J., Adamowicz, W., 2001. The influence of task complexity on consumer choice: a latent class model of decision strategy switching. Journal of Consumer Research 28, 135–148.

Train, K.E., 2001. A comparison of hierarchical Bayes and maximum simulated likelihood for mixed logit. Working Paper, Department of Economics, University of California at Berkeley.

Train, K.E., 2009. Discrete Choice Methods with Simulation. Second Edition, Cambridge University Press, Cambridge.

Tversky, A., 1972a. Elimination by aspects: a theory of choice. Psychological Review 79, 281–299.

Tversky, A., 1972b. Choice by elimination. Journal of Mathematical Psychology 9, 341–367.

Van Cranenburgh, S., Guevara, C.A., Chorus, C.G., 2015. New insights on random regret minimization models. Transportation Research Part A: Policy and Practice 74, 91–109.

Walker, J., Ben-Akiva, M., 2002. Generalized random utility model. Mathematical Social Sciences 43, 303–343.

Williams, H.C.W.L., Ortúzar, J. de D., 1982. Behavioural theories of dispersion and the mis-specification of travel demand models. Transportation Research Part B: Methodological 16, 167-219.

Appendix A. BALANCE IN LATENT CLASS MODELS

We show three examples of the balance of latent classes. In each of these examples, three individuals choose according to class a and three to class b. In all examples, individuals choose the alternatives shown in Table A1.

Table A1. Class of individual and chosen alternative in the balance examples

Individual	Class	Chosen alternative
1	*a*	1
2	*a*	2
3	*a*	2
4	*b*	1
5	*b*	1
6	*b*	2

For simplicity, we assume that the heuristics within each class are identified correctly but the classes used by the individuals are unknown. Therefore, we estimate a LC model with maximum likelihood according to (34) with only one unknown parameter which is the class membership probability π_a .

$$P_{qi}(\pi_a) = \pi_a P_{aqi} + (1 - \pi_a) P_{bqi} \tag{34}$$

Columns 1-4 in Tables A2, A3 and A4 show the probabilities of choosing each alternative when belonging to each class. By changing the probabilities of class b, we manipulate the point of maximum likelihood, which is shown in column 5. Column 6 shows the probability for the chosen alternative given by the latent class model which takes as input π_a and the probabilities of choosing each alternative conditional on the class. Finally, the last column shows the ratio of the probability that each class assigns to the chosen alternative.

Table A2 shows an example where a balance of classes exists. The optimal class membership function indicates that the probability of belonging to class a is 0.31. Note that the balance given by the sum of the ratios of the heuristics and the model has a value equal to the sample size of 6, as stated in Theorem 2 and more generally Theorem 4.

Table A2. *Latent class balanced example*

Heuristic a		Heuristic b		Heuristic a probability	Probability for the chosen alternative		
Alt 1	Alt 2	Alt 1	Alt 2	π_a	P_{q*}	P_{aq*}/P_{q*}	P_{bq*}/P_{q*}
0.50	0.50	0.35	0.65	0.31	0.40	1.26	0.88
0.50	0.50	0.60	0.40	0.31	0.43	1.16	0.93
0.50	0.50	0.70	0.30	0.31	0.36	1.38	0.83
0.50	0.50	0.80	0.20	0.31	0.71	0.71	1.13
0.50	0.50	0.80	0.20	0.31	0.71	0.71	1.13
0.50	0.50	0.30	0.70	0.31	0.64	0.78	1.10
					Sum	6	6

In the second example, shown in Table A3, one of the probabilities –which is underlined– is changed, improving the performance of class b. In this example, the balance still exists but the model estimated probability of belonging to class a decreases. Because the balance still exists, Theorems 2 and 4 hold, showing that the sum of the ratios of the choice heuristic and the models remains equal to the sample size.

Table A3. Latent class low proportion balance example

Heuristic a		Heuristic b		Heuristic a probability	Probability for the chosen alternative		
Alt 1	Alt 2	Alt 1	Alt 2	π_a	P_{q*}	P_{aq*}/P_{q*}	P_{bq*}/P_{q*}
0.50	0.50	0.35	0.65	0.04	0.36	1.40	0.98
0.50	0.50	0.60	0.40	0.04	0.40	1.24	0.99
0.50	0.50	0.64	0.36	0.04	0.37	1.37	0.98
0.50	0.50	0.80	0.20	0.04	0.79	0.63	1.02
0.50	0.50	0.80	0.20	0.04	0.79	0.63	1.02
0.50	0.50	0.30	0.70	0.04	0.69	0.72	1.01
					Sum	6	6

Finally, the third example shown in Table A4 corresponds to a model for which a single latent class is optimal. Even though class a performs better than class b when predicting choices made following class a, the potential benefit of including class a in the model is outweighed by the loss of performance for the last three individuals. Therefore, even though class a is present in the sample, the optimal choice model does not include it. Finally, note that the balance is broken and only the included class ratio sums to the sample size of 6 whilst the excluded class sums to the lower value of 5.89, as is required by Theorem 4.

Table A4. Multiple Heuristic Model example with no balance

Heuristic a		Heuristic b		Heuristic a probability	Probability for the chosen alternative		
Alt 1	Alt 2	Alt 1	Alt 2	π_a	P_{q*}	P_{aq*}/P_{q*}	P_{bq*}/P_{q*}
0.50	0.50	0.35	0.65	0	0.35	1.43	1
0.50	0.50	0.60	0.40	0	0.40	1.25	1
0.50	0.50	0.60	0.40	0	0.40	1.25	1
0.50	0.50	0.80	0.20	0	0.80	0.63	1
0.50	0.50	0.80	0.20	0	0.80	0.63	1
0.50	0.50	0.30	0.70	0	0.70	0.71	1
					Sum	5.89	6

These cases show that the balance can be fragile depending on the probabilities estimated for each heuristic. Even though the underlying process may contain several choice heuristics, a balance among them might not be achieved in estimation.

Appendix B. SIMULATION PARAMETERS

The simulation parameters for each class are given in Table B1, where times are in hours and costs in US$.

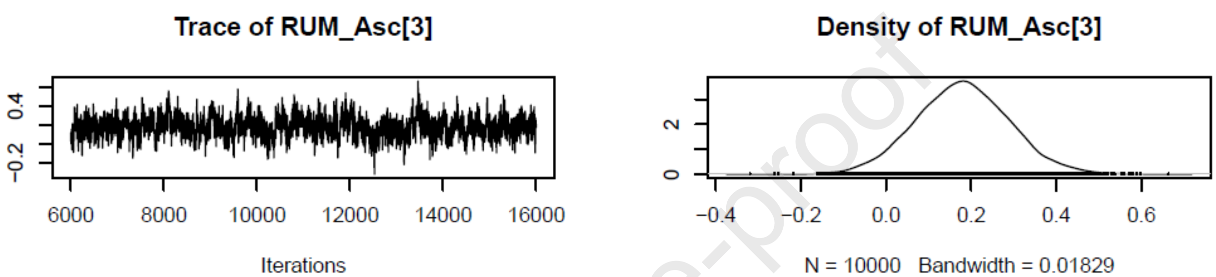
Table B1. Choice heuristic simulation parameters

Parameter	EBA	RRM	SS	RUM
Cost sensitivity	1.39; 1.39	0.375	-6.25	-0.31; +0.09
SS cost threshold	-	-	0.28	-
Vehicle time sensitivity	1.39	2	-12	-5
Waiting time sensitivity	2.30	10	1.5	-20
Walking time sensitivity	2.08	4	4	-6.5
SS time threshold	-	-	0.60	-
μ	-	0.2	-	-
ASC1	0.41	0.1	-0.84	0.5
ASC2	0	0	0	0
ASC3	0.10	0.02	-0.96	0.1
ASC4	0.59	0.16	-0.77	0.8
ASC5	0.53	0.14	-0.80	0.7
ASC6	0.47	0.12	-0.82	0.6
ASC7	0.18	0.04	-0.93	0.2
ASC8	0.26	0.06	-0.90	0.3
ASC9	0.34	0.08	-0.87	0.4

Appendix C. DETECTING IDENTIFIABILITY IN BAYESIAN MODELS

Identifiable models tend to have a stable trace plot as in Figure C1. The trace plot shows a stable mean with no trend and a stable variance, thus exhibiting a clearly-located density function as in the right hand of Figure C1.

Figure C1 Trace plot for identifiable model

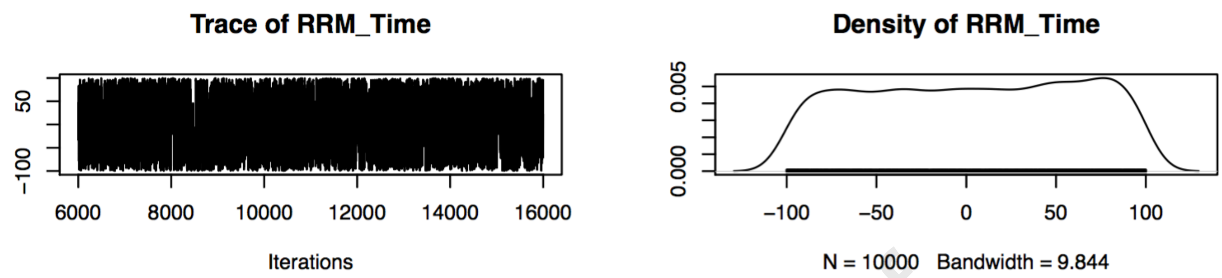


Non-identifiable models tend to have an unstable trace plot as in Figure C2. Under a frequentist approach, a non-identifiable model would have infinite variance given by the inverse of a singular information matrix. Under a Bayesian approach, this is represented by the broad and weakly located posterior distribution.

We adopted the ratio of the standard deviation of the posterior to that of the prior as a quantitative measure of identifiability: small values indicate identifiability whilst values close to one show an absence of information to support identification of the parameter. In the case of the parameter RUM_Asc[3] shown in Figure C1, the prior was a Uniform (-10,10) which has a standard deviation of 5.8 and the posterior distribution had a mean of 0.18 and standard deviation of 0.11, giving a ratio of the standard deviations of approximately 0.019.

In the case of the parameter RRM_Time shown in Figure C2, the prior distribution is Uniform (-100,100), which has a standard deviation of 57.7. The posterior distribution has a mean of 2.5 and a standard deviation of 58.6, giving a ratio of the standard deviations of approximately 1.01 , showing lack of progress After a burn-in of 6,000 samples, a collection of 10,000 samples was used for estimation. Within these, the Markov chain covered the domain of the prior. We assessed this and other similar cases as non-identifiable.

Figure C2 Trace plot for non-identifiable model



Appendix D. COMPARISON OF BAYESIAN AND MAXIMUM LIKELIHOOD ESTIMATION

Table D1 contrasts maximum likelihood (ML) estimation and Bayesian Estimation for two experiments, where the classes were RUM and EBA. Diagnosis of this shows that both estimation procedures achieve strict identifiability. Even though point estimates of the parameters differ, the ML and Bayesian estimates are mutually consistent and are consistent with the values used for simulation

Table D1. Choice heuristic parameters

Experiment		Exp 1. Estimate (Standard deviation)		Exp 2. Estimate (Standard deviation)		
	Parameter	ML	Bayesian	ML	Bayesian	Simulation
Heuristic choice function	θ_0	-0.21 (0.21)	0.07 (0.24)	-0.63 (0.23)	-0.42 (0.26)	0.00
	θ_1	1.42 (0.19)	1.24 (0.18)	1.73 (0.20)	1.57 (0.20)	1.39
RUM	ASC1	0.56 (0.13)	0.55 (0.12)	0.44 (0.13)	0.44 (0.13)	0.50
	ASC 2	0 (fixed)	0 (fixed)	0 (fixed)	0 (fixed)	0.00
	ASC 3	0.17 (0.11)	0.10 (0.11)	0.13 (0.11)	0.13 (0.10)	0.10
	ASC 4	0.86 (0.15)	0.86 (0.15)	0.67 (0.16)	0.63 (0.16)	0.80
	ASC 5	0.74 (0.10)	0.72 (0.10)	0.84 (0.11)	0.83 (0.11)	0.70
	ASC6	0.65 (0.13)	0.66 (0.13)	0.67 (0.13)	0.67 (0.13)	0.60
	ASC 7	0.06 (0.14)	0.06 (0.15)	0.35 (0.14)	0.33 (0.14)	0.20
	ASC 8	0.35 (0.14)	0.35 (0.14)	0.24 (0.14)	0.23 (0.14)	0.30
	ASC 9	0.58 (0.14)	0.56 (0.14)	0.32 (0.15)	0.30 (0.14)	0.40
	Cost	- 0.70 (0.11)	- 0.67 (0.11)	- 0.31 (0.11)	- 0.31 (0.11)	-0.31
	Time:					
	Vehicle	- 5.34 (0.52)	- 5.13 (0.52)	- 5.91 (0.56)	- 5.83 (0.56)	-5.0
	Walk	- 19.9 (1.92)	- 19.8 (1.92)	- 21.4 (2.11)	- 21.0 (2.14)	-20
	Wait	- 7.13 (0.53)	- 6.96 (0.52)	- 6.65 (0.53)	- 6.50 (0.56)	-6.5
EBA	ASC1	0.72 (0.34)	0.70 (0.49)	0.72 (0.34)	0.41 (0.36)	0.41
	ASC 2	1 (fixed)	1 (fixed)	1 (fixed)	1 (fixed)	0
	ASC 3	0.08 (0.40)	-0.37 (0.83)	0.08 (0.40)	-0.70 (0.56)	0.10
	ASC 4	0.82 (0.43)	0.81 (0.62)	0.82 (0.43)	0.70 (0.44)	0.59
	ASC 5	0.60 (0.32)	0.72 (0.46)	0.60 (0.32)	0.14 (0.27)	0.53
	ASC6	0.54 (0.39)	0.41 (0.65)	0.54 (0.39)	0.48 (0.35)	0.47
	ASC 7	0.80 (0.37)	0.92 (0.49)	0.80 (0.37)	-0.38 (0.30)	0.18
	ASC 8	-0.04 (0.39)	-0.14 (0.62)	-0.04 (0.39)	0.32 (0.33)	0.26
	ASC 9	0.46 (0.34)	0.57 (0.46)	0.46 (0.34)	0.26 (0.29)	0.34
	Cost 1	1.04 (0.40)	1.37 (0.67)	1.04 (0.40)	1.35 (0.43)	1.39
	Cost 2	0.70 (0.44)	0.60 (1.10)	0.70 (0.44)	0.65 (0.39)	1.39
	Time:					
	Vehicle	1.67 (0.37)	1.99 (0.63)	1.67 (0.37)	0.81 (0.42)	1.39
	Walk	2.27 (0.38)	2.53 (0.59)	2.27 (0.38)	1.94 (0.45)	2.30
	Wait	2.10 (0.37)	2.42 (0.62)	2.10 (0.37)	1.81 (0.55)	2.08

Appendix E. VERIFICATION OF BALANCE THEOREMS IN BAYESIAN ESTIMATION

We show how the theoretical balance of Theorem 3 holds in the examples simulated. Although it applies strictly to maximum likelihood estimation rather than the Bayesian estimation used here, we show that for our large sample sizes Theorem 3 holds almost exactly. This is consistent with findings in practice (McElreath, 2015; Train, 2001) that Bayesian estimates align closely to maximum likelihood ones as the sample size increases.

For simplicity, we tested the case with no correlation between the class membership function and the choice heuristic parameters. For this case $v = \theta_0 + \theta_1 \cdot trait$ and $\pi_a = \frac{\exp(v)}{1+\exp(v)}$, $\frac{\partial P_{aq*}}{\partial v} = 0$, so according to Corollary 3.1, the balance of Theorem 3 states as (33):

$$\sum_q \frac{\frac{\partial \pi_a(v)}{\partial v} P_{aq*}(\theta)}{P_{q*}(\theta, v)} = \sum_q \frac{\frac{\partial \pi_b(v)}{\partial v} P_{bq*}(\theta)}{P_{q*}(\theta, v)}$$

$$\Rightarrow \sum_q \frac{\pi_a(v) P_{aq*}(\theta)}{(1+\exp(v))\, P_{q*}(\theta, v)} = \sum_q \frac{\pi_b(v) P_{bq*}(\theta)}{(1+\exp(v))\, P_{q*}(\theta, v)} \,. \tag{33}$$

To investigate, in each experiment, whether the sum in (33) for the RUM class has the same value as that for the other class, we calculated their quotient *R*. Results close to 1 show balance between the latent classes, whereas those different from 1 show a lack of balance. Specifically, we calculated expression (34).

$$R = \frac{\sum_q \frac{\pi_a(v) P_{aq*}(\theta)}{(1+\exp(v))\, P_{q*}(\theta, v)}}{\sum_q \frac{\pi_b(v) P_{bq*}(\theta)}{(1+\exp(v))\, P_{q*}(\theta, v)}} \tag{34}$$

In Table 5, non-balance cases show higher instability in the ratio given by (34). Indeed, in the non-balance cases, the standard deviation of the ratio was at least 3-4 times larger than in the cases where a balance was achieved.

Table 5. Theorem 3 verification

Secondary class	Dominant class	Strong Balance cases	Ratio R for balance cases (standard deviation)	Ratio R for non-balance cases (standard deviation)
RRM	RUM	0	-	1.019 (0.043)
	RRM	0	-	0.997 (0.03)
SS	RUM	0	-	1.003 (0.004)
	SS	10	1.001 (<0.001)	-
EBA	RUM	9	1.004 (0.001)	1.005 (-)
	EBA	10	0.997 (0.001)	-

In conclusion, this analysis shows that Theorem 3 applies closely when the model is estimated using Bayesian rather than maximum likelihood methods. Larger standard deviations show instability of the balance. Further work would be required to investigate the rate of approach of Bayesian estimates to the maximum likelihood results of the Theorems.

- we analyse the theoretical properties of latent class models to establish necessary conditions on the classes to be identifiable jointly

- we establish a measure of behavioural difference and relate it to empirical identifiability; this measure highlights factors that are crucial for identifiability

- we provide a graphical diagnostic for identifiability with examples of model non-identifiability, weak identifiability and strong identifiability

Author Statement

Felipe González-Valdés Data curation, Conceptualisation, Investigation, Writing
Benjamin G. Heydecker: Supervision, Validation, Writing-Reviewing and Editing.
Juan de Dios Ortúzar: Conceptualisation, Methodology, Supervision, Validation, Writing-Reviewing and Editing.

We have no conflicts of interest of any form.