



# Modeling safety as a perceptual latent variable to assess cycling infrastructure

Tomás Rossetti<sup>a,b,\*</sup>, C. Angelo Guevara<sup>c</sup>, Patricia Galilea<sup>a,b</sup>, Ricardo Hurtubia<sup>a,b,d</sup>

<sup>a</sup> Department of Transport Engineering and Logistics, Pontificia Universidad Católica de Chile, Chile

<sup>b</sup> Centre for Sustainable Urban Development (CEDEUS), Pontificia Universidad Católica de Chile, Chile

<sup>c</sup> Department of Civil Engineering, Universidad de Chile, Chile

<sup>d</sup> School of Architecture, Pontificia Universidad Católica de Chile, Chile

## ARTICLE INFO

### Keywords:

Discrete choice models  
Integrated choice and latent variable models  
Perceptual latent variables  
Urban cycling

## ABSTRACT

Perceived safety is a relevant construct that affects cyclists' behavior. Nevertheless, this qualitative attribute has not been adequately considered in infrastructure design preference models. Even though the use of perceptual latent variable may be suitable for this, they have not been used, likely because of their high data collection and estimation costs. We propose a change to the classic Integrated Choice and Latent Variables (ICLV) framework that reduces these costs. This methodology is used with data from a stated preferences survey carried out in Santiago (Chile), which collected choice and psychometric indicators. With these, we estimate models with a latent variable that describes perceived safety and latent classes. Results show perceived safety has a positive and relevant effect on infrastructure design preferences. Moreover, heterogeneous preferences are shown to exist, with some users putting a significantly higher weight on perceived safety.

## 1. Introduction

Cities around the world, and particularly in Latin America, have recently invested in cycling infrastructure to promote this mode of transportation. Nevertheless, decision makers are still troubled by the specific design these new bicycle paths or lanes should have. Indeed, in cities like Santiago (Chile) or Bogotá (Colombia) there is no standardized design.

The literature has provided several insights into users' preferences, especially in developed countries. Their main conclusion is that one of the most relevant aspects that should be taken care of is how safe users feel when riding, mainly because users that feel cycling is safer tend to cycle more (Ryley, 2006; Emond et al., 2014; Majumdar Bandhu et al., 2015). Moreover, objective safety measures (i.e., the chances of being involved in a crash) do not explain route choice and the decision to cycle by themselves, but together with perceived safety (Ma and Dill, 2017). Some evidence indicates that these two types of safety do not necessarily correspond. For example, Shankwiler (2006) found that cyclists remember the unsafer sections of routes more than the rest, which means that their subjective valuation of safety can hardly correspond to objective measures of safety. Because of the above, perceived safety is a vital element to be considered when designing and building new infrastructure for bicycles. Nevertheless, to the best of our knowledge, there are no studies that explicitly consider this qualitative construct in the modeling of infrastructure design preferences.

A suitable way of doing this is through the use of latent variables. This approach deals with the explicit modeling of unobserved psychological characteristics or interpretations of decision makers, such as attitudes and perceptions (McFadden and Train, 2000;

\* Corresponding author at: Vicuña Mackenna 4860, Macul, Santiago, Chile.

E-mail address: [terosset@uc.cl](mailto:terosset@uc.cl) (T. Rossetti).

Ben-Akiva et al., 2002). Under this framework, latent variables are constructed as a function of observable ones and simultaneously explain users' choices and responses to psychometric indicators. Unfortunately, perceptual latent variables (those that describe qualitative attributes of alternatives) have hardly been explored in the literature, as opposed to attitudinal latent variables (those that describe qualitative characteristics of users). One of the reasons for this is, probably, the fact that the estimation of perceptual latent variables requires the collection of information on perceived qualitative attributes for each of the alternatives presented in the choice experiments. This is evidently more costly: While attitudinal latent variables demand one indicator per respondent, perceptual ones demand one indicator per alternative presented. This can be very expensive and generate cognitive stress on the respondent. Moreover, these models require the estimation of high-dimensional integrals, which proves to be a major computational challenge in many cases.

To solve this issue, we propose a modified formulation of the integrated choice and latent variable model that heavily reduces surveying and estimation costs. This methodology is used with data collected through a stated preference survey carried out in Santiago (Chile) to estimate choice models that inform on users' preferences for cycling infrastructure design and the impact perceived safety has on these. Results show that the use of latent variables allows for a richer interpretation, especially when combined with latent classes.

What remains of this article is organized as follows: Section 2 presents the general methodology for the estimation of latent variables, along with other relevant discussions involving them. Later, Section 3 presents the modified methodology, which allows an easier estimation of perceptual latent variables. Next, Section 4 presents the case study, along with the models estimated with this data. Finally, Section 5 discusses these results and summarizes this study's main contributions.

## 2. Latent variables in discrete choice models

Following the basic principles of random utility models, one can assume that individual  $n$  has a certain probability of choosing alternative  $i$ , given a vector of observable attributes and characteristics,  $X_{in}$ , a vector of latent variables,  $X_{in}^*$ , a vector of parameters  $\beta$ , and a vector  $\theta_\epsilon$  related to the error term in the utility function behind these probabilities, as seen in Eq. (1).

$$P(y_{in} | X_{in}, X_{in}^*, \beta, \theta_\epsilon) \tag{1}$$

To obtain a probability that does not depend on unobservable characteristics, it is necessary to calculate the expected value of (1). For this, the distribution of the latent variable must be obtained. This distribution can be derived from the latent variable's structural equation, which is shown in (2). Here, vector  $\gamma$  contains parameters to be estimated that relate the latent variable  $X_{in}^*$  with observable attributes and characteristics  $X_{in}$ , and  $\omega_{in}$  is a random error term.

$$X_{in}^* = X^*(X_{in}; \gamma) + \omega_{in} \tag{2}$$

Given an assumption for the distribution of the error term, a density function  $f(X_{in}^* | X_{in}; \gamma, \theta_\omega)$  can be obtained for the latent variable, and the unconditional probability of (1) can be calculated as shown in (3).

$$P(y_{in} | X_{in}; \beta, \gamma, \theta_\epsilon, \theta_\omega) = \int P(y_{in} | X_{in}, X_{in}^*, \beta, \theta_\epsilon) f(X_{in}^* | X_{in}; \gamma, \theta_\omega) dX_{in}^* \tag{3}$$

In this way, a discrete choice model with latent variables could be obtained only observing vector  $X_{in}$  and the choices made by individuals, which is an indicator of these individuals' preferences. Nevertheless, this indicator by itself will unlikely be enough to empirically identify the effects of the estimated latent variables. In other words, the factors that explain the latent variable would not be clear. Therefore, it is useful to include psychometric indicators that also inform on the latent variables of interest (Ben-Akiva et al., 2002). These indicators are included through a measurement equation, as shown in (4), where  $I_{in}$  is the response to the indicator given by individual  $n$  with respect to alternative  $i$ ,  $I$  is a function that relates the latent variable  $X_{in}^*$  with  $I_{in}$  through the use of  $\alpha$ , a vector of parameters. A random error,  $\nu_n$ , is included.

$$I_{in} = I(X_{in}^*, \alpha, \nu_n) \tag{4}$$

With this, a density function  $g(I_{in} | X_{in}^*; \alpha, \theta_\nu)$  can be constructed, which shows with what probability  $I_{in}$  will take a certain value given  $X_{in}^*$ ,  $\alpha$  and  $\theta_\nu$ . In this way, a joint probability can be constructed as shown in (5). Parameters  $\beta$ ,  $\alpha$  and  $\gamma$  can be found maximizing the likelihood for these probabilities. The parameters obtained describe what is usually called Integrated Choice and Latent Variable (ICLV) model.

$$P(y_{in}, I_{in} | X_{in}; \beta, \alpha, \gamma, \theta_\epsilon, \theta_\nu, \theta_\omega) = \int P(y_{in} | X_{in}, X_{in}^*, \beta, \theta_\epsilon) g(I_{in} | X_{in}^*, \alpha, \theta_\nu) f(X_{in}^* | X_{in}; \gamma, \theta_\omega) dX_{in}^* \tag{5}$$

Bahamonde-Birke et al. (2015) present a categorization of latent variables that is useful to understand their use and how they should be modeled. The authors classify latent variables into two groups: attitudinal and perceptual.

### 2.1. Attitudinal latent variables

These are specific to each individual and describe a tendency to act in a particular way based on her experience and temperament (Allport et al., 1935; Pickens, 2005, as cited in Bahamonde-Birke et al., 2015). These latent variables intend to explicitly model attitudes that are not directly measurable, and describe these as a function of observable users' characteristics. This is the most frequent type of latent variable found in the literature. Some examples include concern over the state of the environment (Hess et al.,

2013; Hoyos et al., 2015), risk aversion (Bartczak et al., 2016), concern over privacy (Potoglou et al., 2015), distrust of government and business (Daly et al., 2012) and general disposition towards cooking (O'Neill et al., 2014).

These attitudes are general and do not relate directly to any alternative. Nevertheless, some studies model attitudes that relate specifically to certain alternatives. For example, Hurtubia et al. (2010) modeled a latent variable that describes attitudes against public transport. Even though this attitude is inseparable from this mode of transport, it cannot be described as a function of its attributes. Another example comes from the work done by Yáñez et al. (2010), who modeled perceived reliability, comfort and accessibility in the context of a mode choice model. Even though these are linked to each mode, they are described only as a function of users' characteristics. This means that they are actually attitudes towards certain modes, and do not link observable attributes with qualitative perceptions.

Attitudinal latent variables that are closely linked to alternatives may be used to explain changes in behavior through changes in the sociodemographic composition of a population. Nevertheless, they cannot inform how a certain mode's qualitative attributes could be changed with interventions (for example, increasing the frequency of a certain bus route or the number of seats on subway trains), and if these qualitative attributes affect behavior. Given that policy makers can only change design attributes of alternatives and not the characteristics of a certain population, no major public policy implications can be derived from these models (see Chorus and Kroesen, 2014).

## 2.2. Perceptual latent variables

These variables describe the way individuals process information related to a specific alternative. These variables not only depend on the users' characteristics, but also on the attributes of the alternative. This means that a change in the attributes of a certain alternative will change the value of the latent variable.

Exploration of these types of latent variables have been very scarce in the literature. To the best of our knowledge, only two studies have explicitly treated this issue. One was conducted by Bahamonde-Birke et al. (2015), who exemplified this kind of latent variable in an interurban travel context. In this model, three perceptual variables were estimated: comfort, level of stress and reliability. These variables were not only a function of users' characteristics, but also of the specific attributes of each alternative (number of required transfers, safety level, etc). The other study was the one conducted by Palma et al. (2016), who modeled consumers' preferences for wines. This model included a perceptual latent variable that describes perceived quality as a function of users' knowledge of wine and the alternatives' attributes, such as grape variety and price. A third example of modeling perceptual latent variables, although treated by an alternative approach, is the work of Guevara and Polanco (2016), in which the endogeneity problem that arises due to the omission of latent attributes is addressed using the Multiple Indicator Solution method.

## 3. Modeling perceptual latent variables

Due to the potential perceptual latent variables have in better explaining users' decisions, it seems strange that the literature has hardly explored these types of variables. One reason why this could be happening is because the classic method used to estimate these variables demands having one psychometric indicator per alternative presented. This increases cognitive stress on respondents and data collection becomes more expensive, hindering the use of this technique. Moreover, computational costs are also higher in the case of perceptual latent variables because of the multiple integral of Eq. (5).

This study proposes a way of estimating these variables, using indicators collected separately from the choice experiments. This allows researchers to substantially reduce survey costs, requiring only one indicator per labeled alternative, and in some cases reducing estimation burden, especially when large choice sets are involved. These indicators must be put in the same context as the choice experiments, in such a way that the perceptions inferred through these indicators matches the perceptions experienced during the choice tasks.

The usual ICLV framework, as proposed by Walker and Ben-Akiva (2002) among others, is shown in Fig. 1. In it, rectangular boxes represent observable variables, while ovals represent unobservable or latent ones. Arrows represent relations, where continuous ones represent structural relations and dashed ones represent measurement relations. Red arrows show the structural relation between observable and unobservable variables ( $X_i$ ,  $X_n$  and  $X_{in}^*$  respectively) with the choice model through parameters  $\beta$ , while blue ones show the relation between observable variables ( $X_i$  and  $X_n$ ) with the latent variable ( $X_{in}^*$ ) through parameters  $\gamma$ .

This methodology requires at least one indicator for every alternative available for each individual. In this way, this alternative's attributes affect not only the choice made but also the observed indicator for this particular alternative.

Datasets used to estimate ICLV models usually have some missing psychometric indicators. Instead of discarding these observations, usual practice is to assume the structural relations between observed and latent variables hold for those responses without indicators. Then, latent variables for these cases are simply imputed with the parameters obtained in the estimation process. In other words, the estimation of ICLV models is possible when not every choice is associated with a psychometric indicator by simply constructing the latent variable in these cases with the use of the  $\gamma$  parameters obtained in the estimation. This technique is used when there is missing data, but is not usually used deliberately to reduce surveying costs.

We propose to use this method to reduce surveying and computational costs, specifically when collecting psychometric indicators related to perceptual latent variables, associated with the attributes of alternatives. In this way, we propose to collect a minimum of one indicator per labeled alternative, in an experiment that is framed in the same way as the choices presented, but with an alternative that is not necessarily related to them. Then, instead of collecting at least one psychometric indicator per alternative, this proposal demands at least only one indicator per labeled alternative, taking advantage of the assumption that structural relations of

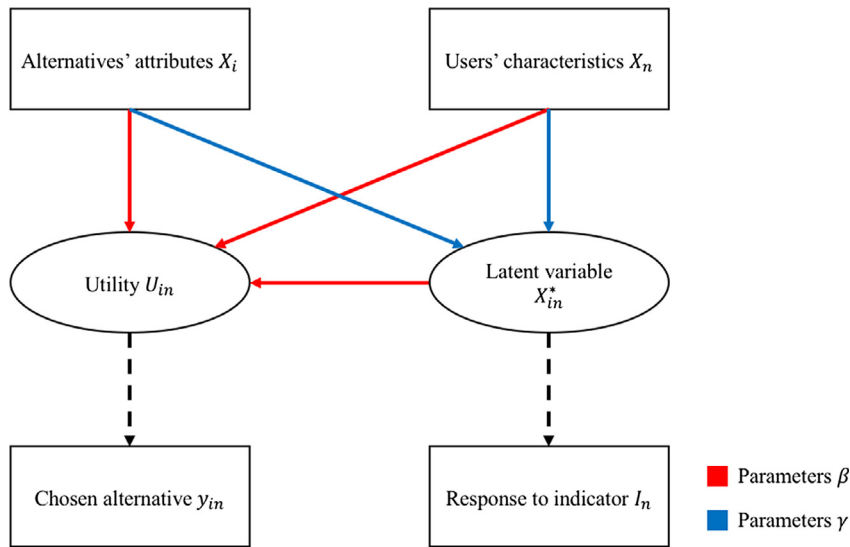


Fig. 1. Framework for a choice model with latent variables (Walker and Ben-Akiva, 2002).

latent variables hold across observations. This allows to reduce surveying and computational costs whenever choice experiments have more than two alternatives, or when respondents are presented with more than one choice experiment, which is usually the case. The number of indicators needed will finally depend on the complexity of the latent variable that the researcher wants to identify, and of its heterogeneity across alternatives and observations. For the linear case usually addressed in practice, a single indicator per labelled alternative would be enough with the proposed method, but more indicators may be recommendable for more complex cases.

The framework for this modified methodology is shown in Fig. 2. This figure explicitly shows that the collected indicator is answered with respect to a certain situation  $j$ , with attributes  $X_j$ , that do not necessarily match the ones from alternative  $i$ . These two models are estimated simultaneously and related to each other through parameters  $\gamma$  from the latent variable model.

With this framework, and following the methodology described previously, Eq. (5) changes. Specifically, the joint probability cannot be expressed as a single integral, but by the multiplication of two. Now, the joint probability is expressed as (6).

$$P(y_{in}, I_{jn} | X_{in}, X_{jn}; \beta, \alpha, \lambda, \theta_\epsilon, \theta_v, \theta_\omega) = P(y_{in} | X_{in}; \beta, \theta_\epsilon) P(I_{jn} | X_{jn}; \alpha, \lambda, \theta_v, \theta_\omega) = \left( \int P(y_{in} | X_{in}, X_{in}^*; \beta, \theta_\epsilon) f(X_{in}^* | X_{in}; \lambda, \theta_\omega) dX_{in}^* \right) \cdot \left( \int g(I_{jn} | X_{jn}^*; \alpha, \theta_v) f(X_{jn}^* | X_{jn}; \lambda, \theta_\omega) dX_{jn}^* \right) \quad (6)$$

Notice that even though the integral in Eq. (6) is separated into two, this probability simultaneously estimates parameters  $\gamma$  for the choice and latent variable models. This sets apart this framework from the sequential modeling approach, ensuring consistent estimators of the model parameters (see Ben-Akiva et al., 2002).

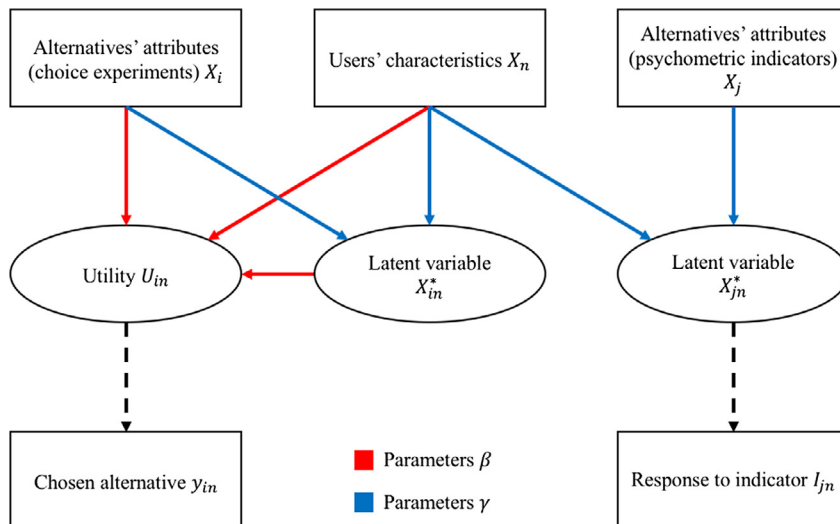


Fig. 2. Framework for a choice model with latent variables used in this study.

For most practical cases, this methodology is more advantageous when respondents face more than one choice task. In this case, and supposing each individual  $n$  answers a total of  $T$  choice experiments, Eq. (6) can be expressed as (7), where  $L_n$  is the combined probability of observing all  $T$  choices correctly.

$$L_n = \int \left( \prod_{t=1}^T P(y_{int} | X_{int}, X_{int}^*; \beta, \theta_\epsilon) f(X_{int}^* | X_{int}; \gamma, \theta_\omega) dX_{int}^* \right) \cdot \int g(I_{jn} | X_{jn}^*; \alpha, \theta_\nu) f(X_{jn}^* | X_{jn}; \gamma, \theta_\omega) dX_{jn}^* \tag{7}$$

Monte Carlo integration is usually used to estimate the parameters  $\beta$ ,  $\alpha$  and  $\gamma$ . Supposing that  $g(I_{jn} | X_{jn}^*; \alpha, \theta_\nu)$  is reduced to a closed-form solution, the above expression can be approximated by (8), where  $X_{intq}^*$  is the  $q$ -th realization of the random vector  $X_{int}^*$  and  $X_{jnr}^*$  is the  $r$ -th realization of the random vector  $X_{jn}^*$ . Notice that this method also has the potential to reduce computational costs by separating these integrals.

$$L_n \approx \left( \prod_{t=1}^T \left( \frac{1}{Q} \sum_{q=1}^Q P(y_{int} | X_{int}, X_{intq}^*; \beta, \theta_\epsilon) \right) \right) \cdot \left( \frac{1}{R} \sum_{r=1}^R P(I_{jn} | X_{jnr}^*; \alpha, \gamma, \theta_\nu, \theta_\omega) \right) \tag{8}$$

#### 4. Case study

This section presents an application of the proposed methodology for the estimation of perceptual latent variables. The application deals with the preferences for cycling infrastructure design, measured through a stated preferences survey conducted in Santiago (Chile). Section 4.1 describes details of the survey, while Section 4.2 presents the results for the estimation of several models with this methodology.

##### 4.1. Data used

During the past years, the city of Santiago (Chile) has seen an important increase in bicycle usage. While during 2001 only 1.9% of all trips during a workday were carried out on these types of vehicles, in 2012 this mode share doubled to 4.0% according to the latest travel survey available (SECTRA, 2015). This increase could be explained partly because new cycling infrastructure has been built in several streets of Santiago during the last decade. Nevertheless, in many cases this infrastructure has been poorly designed (Vega et al., 2015). Moreover, many cyclists today decide to ride on sidewalks, breaking Chilean law and reducing pedestrians' comfort. These users generally argue that they do not feel safe riding on the road.

Understanding how design attributes of bike lanes may induce different riding behavior on cyclists is necessary to generate better infrastructure. Moreover, if sidewalks are perceived as safer, it is necessary to understand to what extent specific design elements on road-level cycling facilities can compensate for the perceived safety of riding at the sidewalk.

This study uses data collected by a stated preferences (SP) survey carried out during March 2016 that aims to understand how cyclists choose streets to ride on. The survey was mainly applied through interviews carried out on busy bike lanes or bike paths and on web-based questionnaires distributed through social media. 1,966 surveys were successfully responded, completing 10,223 choice scenarios. 62% of the sample was male, which is in line with a gender imbalance present among Santiago's cyclists (SECTRA, 2013; Vega et al., 2015). Most respondents were between 20 and 30 years old, showing the sample over-represents young people. Moreover, more than half stated a monthly household income of over 1,500 USD, which could be considered to be of the upper income groups in Santiago. Finally, most respondents stated using a bicycle in their last commute. Details on the deployment of the survey are described in Rossetti (2017).

Five variables were selected to be included in the experiments, at different levels each, for unlabeled, dichotomous experiments:

1. Type of cycling infrastructure present: This variable assessed the general design of the cycling infrastructure, and was presented at five levels.
  - (a) No cycling infrastructure present. Cyclists should ride on the street. An example is shown in Fig. 3a.
  - (b) Bicycle lane present, only marked by a line on the pavement. An example is shown in Fig. 3b.
  - (c) Bicycle lane present, marked by a line and painted surface. An example is shown in Fig. 3c.
  - (d) Bicycle lane present, marked by a line and protected with physical barriers that prevent cars from invading it. An example is shown in Fig. 3d.
  - (e) Bicycle path present, built at the sidewalk level. An example is shown in Fig. 3e.
2. Width: This variable assessed if cyclists prefer to ride on wider infrastructure. It included two levels; 1.5 and 3 meters (5 and 10 feet).
3. Motorized traffic speed: This variable was presented linked to the speed limit allowed on the street. It included two levels: 30 km/h (19 mph) and 60 km/h (37 mph).
4. Presence of public transport buses: This variable included two levels: no buses on the street, or presence of some transit buses.
5. Travel time: This variable was set at three levels: 15, 20 and 30 min.

The experiments were presented with the use of images to control in a better way what respondents were considering, and also because they have the potential to increase the efficiency of parameters (see Hoffer, 2015; Hurtubia et al., 2015). Some examples of

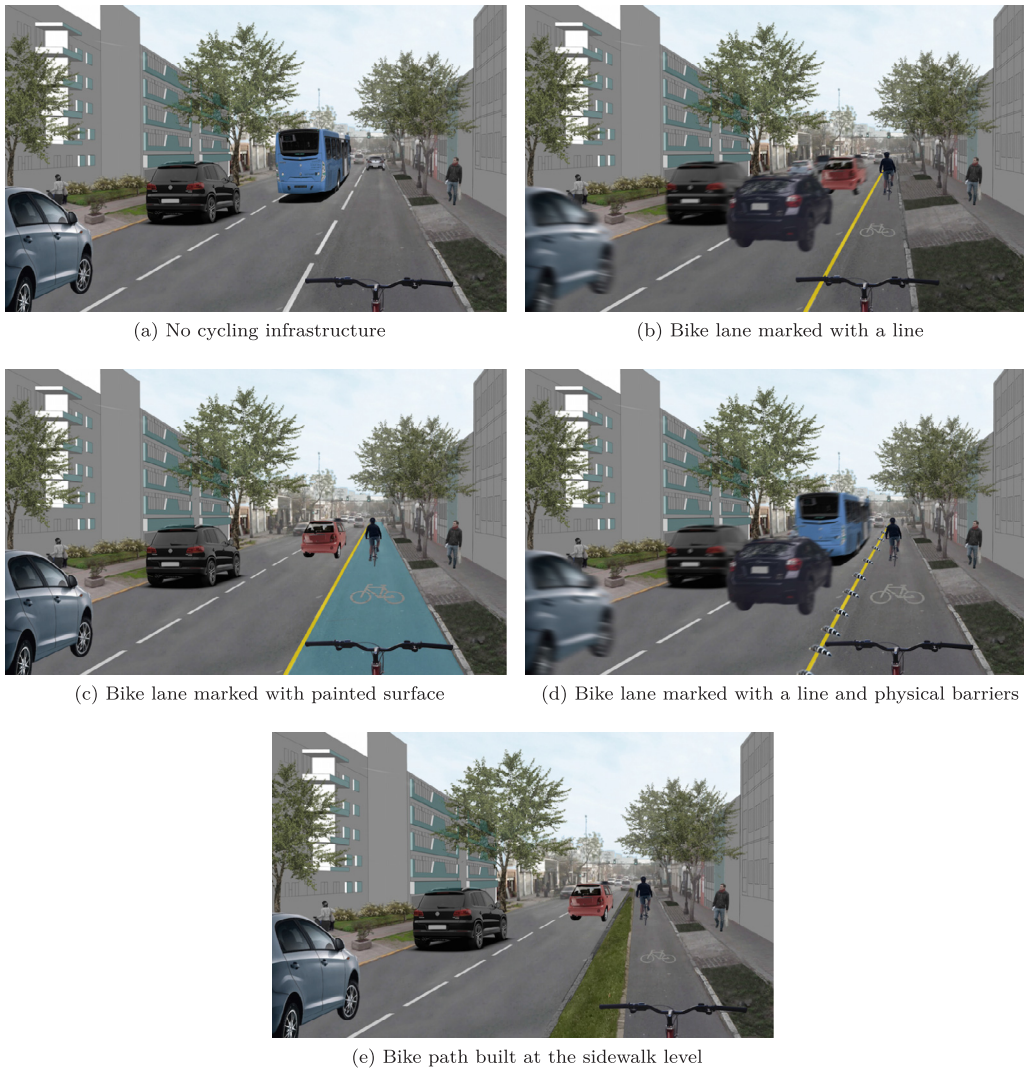


Fig. 3. Examples of images used in the stated preferences experiments, classified by type of cycling infrastructure.

these are shown in Fig. 3, and an example of an experiment presented is shown in Fig. 4. These experiments were constructed using a D-efficient design (ChoiceMetrics, 2012), grouped in three blocks with four questions each.

After the SP tasks were presented to respondents, a random image used in the experiments was shown. Considering this image, they were asked to rate how safe and comfortable they thought this particular street is, and in general how much they like it. An example of this question is presented in Fig. 5. The answer to the question related to safety is the psychometric indicator used in the estimation of ICLV models presented in the next section.

#### 4.2. Results

This section shows the estimation results for two types of models: mixed logit models (MixL) and latent class mixed logit models (LCMixL). For every type of model, two specifications are compared: one without a perceptual latent variable, and one with a perceptual latent variable that models perceived safety. All models were estimated with the use of Biogeme (Bierlaire, 2003).


We used a mixed logit specification to account for the panel nature of the data. The parameter related to the presence of buses on the road was kept constant across all answers given by the same respondent, and each model includes a standard deviation for this parameter among the surveyed population. For further information on mixed logit models, see McFadden and Train (2000).


##### 4.2.1. Mixed logit models (MixL)

Two specifications are presented for the MixL family of models. The first one is a binomial mixed logit model, while the second is an integrated choice and latent variable (ICLV) model. These specifications, shown in Table 1, will be described sequentially in the following paragraphs.

Which route do you prefer?


**Option A**




**60 km/h**  30 min.

**Option A**  
 Trip is 30 minutes long  
 Cars are fast (60 km/h)  
 No public transport buses

**Option B**



**30 km/h**  15 min.

**Option B**  
 Trip is 15 minutes long  
 Cars are slow (30 km/h)  
 Public transport buses present

Fig. 4. Example of stated preferences experiment presented.



**60 km/h**  20 min.

State how much you agree with the following statements regarding the image presented above with a grade from 1 to 7, where 1 is "I strongly disagree" and 7 is "I strongly agree."

I strongly disagree		I disagree		Neither agree nor disagree		I agree		I strongly agree
1		2		3		4		5
		6		7				

This route is safe

● \_\_\_\_\_

This route is comfortable

● \_\_\_\_\_

I like this route

● \_\_\_\_\_

Fig. 5. Example of questions surveying psychometric indicators regarding qualitative attributes.

The first model, which does not include latent variables, allows to understand the main effects each attribute has on respondents' choices. It is the simplest model, and does not include the information contained by the perceptual indicators in its likelihood function. This choice model's utility was modeled linearly, with no interaction between variables. Furthermore, all variables are dummies with the exception of travel time. Due to the fact that there might be a bias towards alternatives presented on the left or

**Table 1**  
Results for the estimation of three MixL models.

	Variable	MixL Value (t-ratio)	ICLV Value (t-ratio)
Choice	Constant (right-hand side)	-0.105** (-4.48)	0.0198 (0.26)
	Bike lane: None	0 (fixed)	0 (fixed)
	Bike lane: Line	0.824** (13.94)	-2.19** (-3.46)
	Bike lane: Painted	1.08** (16.94)	-2.41** (-3.22)
	Bike lane: Protected	1.19** (20.16)	-2.59** (-3.54)
	Bike lane: Sidewalk	1.47** (27.74)	-2.92** (-3.64)
	Wide	0.233** (6.28)	-0.679** (-3.11)
	Low max. speed	0.479** (16.97)	
	No buses (mean)	0.662** (18.38)	-0.0772 (-0.42)
	No buses (std. dev.)	0.811** (19.82)	0.901** (16.93)
	Travel time (minutes)	-0.0682** (-16.83)	-0.0896** (-9.50)
	Perceived safety		7.80** (5.42)
	Perceived safety (Latent var.)	Bike lane: None	
Bike lane: Line			1.22** (7.52)
Bike lane: Painted			1.54** (10.09)
Bike lane: Protected			1.74** (12.12)
Bike lane: Sidewalk			2.11** (13.34)
Wide			0.627** (7.01)
Low max. speed			0.314** (7.02)
No buses			0.432** (5.02)
LV error scale (left-hand side)			0.940** (22.85)
LV error scale (right-hand side)			1 (fixed)
		<i>log-likelihood</i>	-6,138.47
	<i>log-likelihood (choice model)</i> <sup>a</sup>	-6,138.47	-6,131.79
	<i>Number of observations</i>	10,223	10,223
	<i>Number of parameters (choice model)</i>	10	11
	<i>Adjusted rho-squared (choice model)</i>	0.132	0.133

\*\*  $p < 0.01$ ; \*  $p < 0.05$ .

<sup>a</sup> This likelihood is equal to the sum of the logarithm of all choice probabilities.

right-hand side of the question, a dummy was included to control this effect. Results for these models are shown in Table 1, column “MixL.”

All parameters obtained in this model show an effect that is consistent with what the literature has concluded, whether through stated (Stinson and Bhat, 2003; Tilahun et al., 2007; Sener et al., 2009; Caulfield et al., 2012; Evans-Cowley and Akar, 2014; Motoaki and Daziano, 2015; Mertens et al., 2016; Rossetti et al., 2017) or revealed preferences methods (Howard and Burns, 2001; Dill and Gliebe, 2008; Menghini et al., 2010; Hood et al., 2011; Larsen and El-Geneidy, 2011; Broach et al., 2012; Kang and Fricker, 2013; González et al., 2016). Users showed preference for cycling infrastructure with a greater separation from traffic; that is, results show that, *ceteris paribus*, respondents prefer bike paths next to sidewalks the greatest, followed by protected bike lanes, bike lanes with painted surface and bike lanes marked with a separating line. Moreover, the presence of any kind of cycling infrastructure seems to impact users’ choices in a greater manner than any of the other variables included.

Results also show that respondents prefer wider infrastructure, and streets with low maximum speeds and no buses present. Even though the first two variables have been tested in the literature, the presence of public transport buses on the road has not been tested enough; only one study addressed this issue, and in the case of Santiago as well (Hurtubia et al., 2015).

The second model estimated in this family is a MixL that explicitly models perceived safety with the use of a latent variable, and includes this latent variable in the choice sub-model logarithmically. Several specifications were tested, with this one providing a better fit to the data. This means that perceived safety is better represented having a decreasing effect on overall choice; in other words, this implies that an increase in perceived safety has a greater effect on choice when an alternative is deemed as unsafe. Results are shown in Table 1, column “ICLV”.

The latent variable was modeled through the structural equation shown in Eq. (9), where  $X_i^*$  is the perceived safety of alternative  $i$ , and parameters  $\gamma$  relate perceived safety linearly with observable attributes,  $X_i$ . This latent variable was then included logarithmically in the utility function, as shown in Eq. (10). Because the argument of the logarithm has to be strictly positive, the error term  $\omega_i$  was modeled to have a log-normal distribution.

$$X_i^* = \sum_k \gamma_k X_{ik} + \omega_i \tag{9}$$

$$U_i = \sum_k \beta_k X_{ik} + \beta^* \ln(X_i^* + 1) \tag{10}$$



The measurement equation was constructed in such a way that the ordered nature of the indicators was accounted for, following recommendations found in the literature (Daly et al., 2012; Atasoy et al., 2013). The error term associated with it was supposed to follow a Normal distribution, which results in an ordered probit model. Nevertheless, this was approximated through an ordered logit model (McCullagh, 1980), which has a closed-form solution and negligible discrepancies with the ordered probit one (Lee, 1982; Ruud, 1983). This measurement equation was modeled in such a way that the latent variable explains the choice for the indicator shown in Fig. 5; this is, the indicator used describes how safe the respondent would feel riding on the street shown in that specific question. This allows to infer that  $X_i^*$ , the latent variable, can be called “Perceived safety.”

Results from the latent variable sub-model show that users tend to feel safer riding on streets with cycling infrastructure, where this infrastructure is wide, where speed limits are low and where there are no public transport buses present. Respondents stated higher levels of perceived safety on infrastructure with the greatest separation from traffic; this is, they reported they would feel the safest cycling at the sidewalk level, followed by a protected bike lane. These findings are in line with what has been concluded earlier in different studies (Klobucar and Fricker, 2007; Parkin et al., 2007; Chataway et al., 2014; Dill et al., 2015).

The interpretation of the results in the choice sub-model are not as straightforward as the ones in the latent variable one. In the first place, the parameter related to the latent variable “Perceived Safety” has a positive sign, indicating that users prefer to ride on streets that they perceive to be safer. This finding, although intuitive, has not been well established earlier in the literature. What has usually been done is that researchers interpret the results of choice models without latent variables and presume that safety is a main driver of these (see, for example, Sener et al., 2009). Nevertheless, a link has not been explicitly established.

The other parameters represent effects that haven’t been captured by “Perceived Safety” but that have a significant effect on users’ decisions. For example, negative parameters related to all types of bike lanes indicate that there are other latent factors that affect these parameters, such as comfort, beauty, the ability to relax or perceived speediness. The attribute related to the presence of buses on the road is non-significant, which means that all of the effect of this attribute on choice is explained by perceived safety.

It is important to note that, in the model including latent variable, the “Low max. speed” variable is not present in the utility function, but is included in the latent variable’s structural equation. This was the best specification found that includes this variable.

The ICLV model’s choice likelihood is significantly higher than the one without latent variables.<sup>1</sup> It is important to note that, as Vij and Walker (2016) prove, this does not mean that the use of psychometric indicators is responsible for this increase. Indeed, it is possible that a non-linear model with a similar specification may have a similar fit. The use of latent variables does not help by itself to achieve better goodness-of-fit measures, but they may useful explore specifications that may have a better fit and retain proper causality relations.

The overall effects of the ICLV model are not straightforward to obtain because the effect each variable has on the overall choice is affected partly by the choice sub-model and by the latent variable sub-model. One way of measuring these effects is by detecting the marginal effect each variable has on choice; in other words, these effects can be detected by calculating the partial derivatives for the utility functions. The mean utilities for the models presented above,  $\bar{U}_i^{\text{MixL}}$  and  $\bar{U}_i^{\text{ICLV}}$ , are shown in Eqs. (11) and (12), where  $\hat{\beta}$  and  $\hat{\gamma}$  are the mean maximum likelihood estimators for parameters  $\beta$  and  $\gamma$ .

$$\bar{U}_i^{\text{MixL}} = \sum_k \hat{\beta}_k X_{ik} \tag{11}$$

$$\bar{U}_i^{\text{ICLV}} = \sum_k \hat{\beta}_k X_{ik} + \hat{\beta}^* \ln \left( \sum_k \hat{\gamma}_k X_{ik} + 1 \right) \tag{12}$$

The total effect a certain variable,  $X_{ik'}$ , has on these utilities can be obtained simply calculating their derivatives with respect to this specific variable. The derivatives of the utility functions above are shown on (13) and (14).

$$\frac{\partial \bar{U}_i^{\text{MixL}}}{\partial X_{ik'}} = \hat{\beta}_{k'} \tag{13}$$

$$\frac{\partial \bar{U}_i^{\text{ICLV}}}{\partial X_{ik'}} = \hat{\beta}_{k'} + \frac{\hat{\beta}^* \hat{\lambda}_{k'}}{\sum_k \hat{\lambda}_k X_{ik} + 1} \tag{14}$$

While the first cases’ values are independent from the attributes of the alternative, the second case is a function of  $X_i$ . This means that the net effect of each variable is dependent on the specific configuration of the alternative. This is the case when latent variables are included in the utility functions in non-linear ways. The net effects for the MixL models (described in detail in Table 1) are shown in Table 2.

#### 4.2.2. Latent class mixed logit models

Studies that have focused on perceived safety on cyclists have discovered that users tend to have heterogeneous preferences. Specifically, many have concluded that women tend to feel less safe than men while riding a bicycle (Bernhoft and Carstensen, 2008; Chataway et al., 2014; Dill et al., 2015). This gender disparity has also been noted more generally on route choice while cycling (Aldred et al., 2016). Experience riding bicycles has also been proved to be a source of heterogeneity while choosing a road to cycle

<sup>1</sup> Goodness-of-fit comparison test between two non-nested models (Horowitz, 1983):  $P(\hat{\beta}_2^2 - \hat{\beta}_1^2 > z) \leq 0.000222$ .

**Table 2**  
Net effects of MixL models.

Variable	No LV	Log LV <sup>a</sup>		
		Min.	Average	Max
Constant (right-hand side)	−0.105		0.0198	
Bike lane: None	0		0	
Bike lane: Line	0.824	0.458	0.966	1.40
Bike lane: Painted	1.08	0.660	1.30	2.32
Bike lane: Protected	1.19	0.983	1.50	1.85
Bike lane: Sidewalk	1.47	1.03	1.42	2.37
Wide	0.233	0.494	0.627	0.868
Low max. speed	0.479	0.605	0.776	1.86
No buses (mean)	0.662	0.731	0.936	2.28
No buses (std. dev.)	0.811		0.901	
Travel time	−0.0682		−0.0896	
<i>log-likelihood (choice model)</i>	<i>−6,138.47</i>		<i>−6,131.79</i>	

<sup>a</sup> The effect of the “Log LV” model is a function of the attributes of the alternatives. Therefore, a single value cannot be provided. The “Min.” and “Max.” columns show the minimum and maximum possible values of  $\partial \bar{U}_i^{\text{CLV}} / \partial X_{ik}$  given the alternatives showed to respondents, and “Average” is the average value of the function considering the 20,446 alternatives presented to all users.

on. Many studies have concluded that users that cycle more often tend to prefer faster routes and with less separation from traffic (Heinen et al., 2010). Finally, age has been found to be relevant in some studies; generally, evidence shows that older people tend to prefer routes that are more separated from vehicular traffic. Nevertheless, this evidence is weaker (Aldred et al., 2016). Failure to account for this heterogeneity could result in biased estimators (Fowkes and Wardman, 1988; Ortúzar and Willumsen, 2011).

To consider this preference diversity, latent classes were built upon the models presented in the previous section. These models segment the population into a discrete number of groups with relatively homogeneous preferences. They have been explored in the literature in many scenarios, such as residential location (Walker and Li, 2007), demand for cultural amenities (Grisolia and Willis, 2012) or mode choice (Bhat, 1997; Shen, 2009; Hurtubia et al., 2014), reaching better fit to the data and a richer interpretability.

Choice models with latent classes have two sub-models: one which describes membership to a certain class, and one specific choice model for each class created. The class-membership model supposes individuals do not belong to a certain class deterministically, but stochastically. This means that each person belongs to a latent class with a given probability, and therefore cannot be classified into one or another class with certainty (see Walker and Ben-Akiva, 2002).

With the data available, a latent class model was added to the models presented above. In this specific case, two latent classes were built. In order to estimate these, and following notation used by Walker and Ben-Akiva (2002), two class-specific utility functions were estimated. The one related to Class 1,  $U_{1n}$ , shown in (15), was a function of users’ attributes  $X_n$ , and related them to the utility linearly through parameters  $\lambda$ . On the other hand, the deterministic part of utility function specific to Class 2,  $U_{2n}$ , shown in (16), was set to equal zero.

$$U_{1n} = W(X_n; \lambda) + \zeta_{1n} \quad (15)$$

$$U_{2n} = \zeta_{2n} \quad (16)$$

Error terms  $\zeta_{1n}$  and  $\zeta_{2n}$  were assumed to be independent and identically distributed Extreme Value Type I, which means that class membership probabilities follow a binary logit form. The latent class sub-model was constructed with five users’ characteristics, which are defined as follows:

- “Female”: dummy variable equal to one if the user is a woman, zero otherwise.
- “Experienced”: dummy variable equal to one if the user has been using a bicycle to commute for a year or more, zero if the user commutes by bicycle for less time or does not use a bicycle for her daily commute.
- “Age”: continuous variable, equal to the user’s stated age.
- “Public bicycle”: dummy variable equal to one if the user is subscribed to one of Santiago’s bike-sharing systems, zero otherwise.
- “High income”: dummy variable equal to one if the user stated a household income equal to or over CLP 1,000,000 (USD 1,500 approximately), zero otherwise.

These models were estimated in a similar fashion as in the previous section: one did not include a latent variable, the second one includes it linearly, and the third one includes it logarithmically. Utilities had the same specifications as the ones shown above. Results for this family of models are shown in Table 3.

Both models presented similar latent class models. In these, users showed a greater probability of belonging to Class 1 if they were experienced cyclists, male, young, did not use public bicycles and belonged to higher income groups. Considering previous evidence in other contexts, this class seems to group regular cyclists who have a low fear of traffic, and will therefore be called “Fearless”. On the other hand, results show users belonged to Class 2 with a greater probability if they were inexperienced, older, lower-income women who did not cycle to work. This class is the opposite of Class 1: people that tend to fear cycling and prefer to ride on

**Table 3**  
Results for the estimation of three latent class MixL models.

		Variable	MixL Value (t-ratio)	ICLV Value (t-ratio)
Latent class	Class 1	Constant	0.00633 (0.02)	-0.0155 (-0.06)
		Female	-0.338* (-2.25)	-0.310* (-2.07)
		Experienced	0.764** (5.52)	0.761** (5.46)
		Age	-0.0271** (-3.90)	-0.0275** (-3.95)
		Public bicycle	-0.629* (-2.51)	-0.634* (-2.52)
		High income	0.289* (2.08)	0.296* (2.12)
Choice	Class 1 (“Fearless”)	Constant (right-hand side)	0.125 (1.82)	0.141* (2.11)
		Bike lane: None	0 (fixed)	0 (fixed)
		Bike lane: Line	1.35** (7.68)	1.331** (4.45)
		Bike lane: Painted	1.62** (11.59)	1.59** (5.36)
		Bike lane: Protected	1.44** (11.10)	1.40** (4.21)
		Bike lane: Sidewalk	0.711** (6.41)	0.683* (2.05)
		Wide	-0.112 (-1.24)	-0.149 (-1.55)
		Low max. speed	0.0101 (0.15)	
		No buses (mean)	0.564** (5.20)	0.570** (4.01)
		No buses (std. dev.)	0.274 (1.23)	0.304 (1.36)
	Class 2 (“Cautious”)	Travel time	-0.159** (-9.57)	-0.162** (-9.71)
		Perceived safety		0.0697 (0.14)
			-0.151** (-3.10)	0.0445 (0.46)
		Bike lane: None	0 (fixed)	0 (fixed)
		Bike lane: Line	0.830** (7.82)	-1.74** (-3.47)
		Bike lane: Painted	0.850** (6.83)	-2.21** (-3.80)
		Bike lane: Protected	1.22** (10.89)	-2.06** (-3.87)
		Bike lane: Sidewalk	2.19** (15.11)	-1.64** (-2.94)
		Wide	0.396** (5.37)	-0.596* (-2.54)
		Low max. speed	0.896** (11.12)	
Perceived safety (Latent var.)	No buses (mean)	0.976** (9.10)	0.106 (0.59)	
	No buses (std. dev.)	1.22** (13.31)	1.29** (11.47)	
	Travel time	-0.0319** (-4.12)	-0.0420** (-3.97)	
	Perceived safety		7.73** (6.37)	
	Bike lane: None		0 (fixed)	
	Bike lane: Line		1.13** (6.87)	
	Bike lane: Painted		1.38** (8.52)	
	Bike lane: Protected		1.58** (10.53)	
	Bike lane: Sidewalk		1.94** (11.57)	
	Wide		0.671** (7.41)	
Low max. speed		0.534** (9.00)		
No buses		0.498** (5.73)		
Error std. dev. (left-hand side)		0.912** (17.82)		
Error std. dev. (right-hand side)		1 (fixed)		
<i>log-likelihood</i>			-5,859.78	-9,139.93
<i>log-likelihood (choice model)<sup>a</sup></i>			-5,859.78	-5,860.99
<i>Number of observations</i>			10,223	10,223
<i>Number of parameters (choice model)</i>			26	26
<i>Adjusted rho-squared (choice model)</i>			0.169	0.169

\*\*  $p < 0.01$ ; \*  $p < 0.05$

<sup>a</sup> This likelihood is equal to the sum of the logarithm of all choice probabilities.

infrastructure with high levels of segregation. Therefore, this class will be referred to as “Cautious”.

Fig. 6 shows a histogram for the predicted probabilities of belonging to both latent classes among the surveyed population. It can be seen that most individuals have a greater probability of belonging to Class 2, which describes “Cautious” cyclists. A minor amount of individuals have a probability of belonging to Class 1 greater than 0.5, and none greater than 0.7. This means that the results for the Fearless’ choice model should be read with caution; no individuals are described perfectly by this class, but rather have a higher or lower composition of this archetype in their choice patterns.

The MixL model with latent classes shows preferences among the two groups modeled show different behavioral conclusions. For example, the Fearless group shows non-significant parameters associated with infrastructure width and maximum speeds on streets. On the other hand, the Cautious latent class shows significant parameters associated with these variables, with expected signs.

Infrastructure design also showed differences in preferences: while Cautious users tended to prefer infrastructure at the sidewalk level, this option was Fearless’ least preferred design. These users showed a greater preference for infrastructure at the road level,

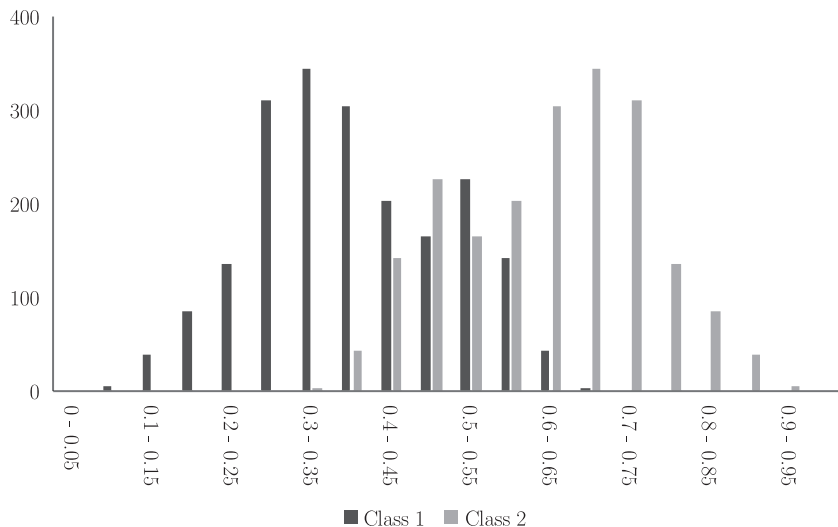


Fig. 6. Histogram for the estimated class-membership probabilities for the LCMixL model with no latent variables.

with no significant differences between these three types of design.<sup>2</sup> Results, therefore, allow to state that Cautious users greatly prefer cycling at the sidewalk level, while Fearless users prefer to ride at the road level, with no particularly preferred design.

The ICLV model's most interesting result is that Fearless users' choice sub-model presented a non-significant parameter related to perceived safety. On the other hand, Cautious users presented a positive and significant parameter related to this latent variable. Because of the way the probabilities were distributed among the individuals, it is incorrect to state that there are users that do not consider safety while cycling. Nevertheless, these results show that users value safety in different ways, and that those users that are better described by the Fearless archetype have a lower concern about safety. Note that this conclusion can be reached because the latent variables were included in the class-specific choice models, which is an uncommon practice.

Table 3 shows that the MixL and ICLV models with latent classes have goodness-of-fit measures that do not differ significantly, in spite of the increase the "Log LV" model showed in the previous section. This means that at an aggregate level perceived safety is best modeled with a decreasing effect on security, but in this disaggregate scenario there might be better ways of representing it.

## 5. Conclusions

Hybrid discrete choice models have been heavily used during the past years to explicitly model latent constructs that affect decision patterns. Nevertheless, this exploration has had an emphasis on attitudinal latent variables (those that describe users' latent characteristics), leaving perceptual latent variables (those that describe alternatives' latent attributes) mostly untouched. This could be due to the fact that the current methodology recommended for the estimation of integrated choice and latent variable models require an important amount of information, which increases surveying costs.

This paper proposes a modification of this methodology in order to lower these costs. Instead of collecting one psychometric indicator for each alternative presented to respondents, this paper proposes a methodology that allows to collect only one indicator per labeled alternative and respondent. With these indicators, along with the observed choices, integrated choice and latent variable models can be estimated.

This proposal was formulated to consider the effect perceived safety has on preferences for bicycle infrastructure design. Data from a stated preferences survey was used. In it, respondents were presented with different hypothetical streets on which to cycle, and were asked to show their preferences. After this, users were presented with a random alternative, and asked to state how safe they would feel cycling on that particular street. With this data, two types of models were estimated: mixed logit ones, and mixed logit ones with latent classes. Both types included a latent variables sub-model that explicitly modeled perceived safety.

Both types of models showed that the use of perceptual latent variables was useful to better understand users' decisions. Specifically, perceived safety was shown to be significant in this choice context. This relation, although expected, has not been explicitly established before in the literature. The mixed logit models with latent classes also showed that certain users place a far smaller weight on perceived safety when deciding where to ride. These users tend to be experienced, younger men from higher income groups who do not ride on public bicycles.

The identification of two distinct types of users shows there is a relevant paradox that has to be considered in the policy cycle. Even though those who cycle prefer to do it at the road level, those who do not cycle would prefer to do so at the sidewalk level. The

<sup>2</sup> Difference between "Line" and "Painted" infrastructure: z-test equal to 1.84. Difference between "Line" and "Protected" infrastructure: z-test equal to 0.63. Difference between "Painted" and "Protected" infrastructure: z-test equal to 1.45. All tests are under the critical value of 1.96 for a two-tailed test with a 95% significance level.

decision of which type of infrastructure to build is not obvious. If policy makers decide to build at the street level, current users will be content, but some potential cyclists will not feel safe enough to start using this vehicle. On the other hand, if decision makers build facilities at the sidewalk level, new users will be attracted, but current users will not be satisfied. Moreover, it is possible that new users will change their preferences and, in the long run, also prefer infrastructure at the road level.

We believe this question is still open for debate. Nevertheless, we see two possible paths of action. One is to build segregated infrastructure at the road level, which is preferred the most by experienced users and liked in a lesser degree by inexperienced users. This type of infrastructure is also perceived to be safe by inexperienced users, which is why it could have the potential to attract new cyclists. On the other hand, a possible path of action is to build different types of infrastructure in coexisting networks. This could provide different types of users with routes more suited to their preferences and needs, but could potentially be more costly to local governments. Moreover, the evidence states that there is an increased risk of injury while riding on sidewalks (Wachtel and Lewiston, 1996; Moritz, 1998; Aultman-Hall and Kaltenecker, 1999; Reynolds et al., 2009), which could mean that sidewalk-level facilities may be less desirable.

Our final recommendation for policy makers is to build infrastructure where it is most needed: busy streets, with fast traffic and with presence of public transport buses. If it is too costly, then city officials must make sure there are alternatives nearby with enough capacity. Investment on streets with low maximum speeds and low volumes of traffic will likely be unprofitable given the needs of most cities around the world for more cycling infrastructure. We recommend focalizing investment on those streets where cycling is less attractive.

Understanding how cyclists perceive a city's streets is vital to better focalize public investment on infrastructure. The methodology proposed in this study could be replicated for other qualitative attributes of streets, such as beauty or comfort, that could better guide infrastructure design and public policy in general.

## Acknowledgments

The authors gratefully acknowledge the research support provided by the Centro de Desarrollo Urbano Sustentable (Center for Sustainable Urban Development, CEDEUS, CONICYT/FONDAP 15110020), the Instituto de Sistemas Complejos de Ingeniería (Institute of Complex Engineering Systems, ISCI, ICM: P-05-004-F, CONICYT: FBO16) and CONICYT (FONDECYT 1150590, FONDECYT 11130637).

## References

- Aldred, R., Elliott, B., Woodcock, J., Goodman, A., 2016. Cycling provision separated from motor traffic: a systematic review exploring whether stated preferences vary by gender and age. *Transp. Rev.* 1647 (July), 1–27.
- Allport, G.W., 1935. Attitudes. In: *A Handbook of Social Psychology*, vol. 1, pp 798–844.
- Atasoy, B., Glerum, A., Bierlaire, M., 2013. Attitudes towards mode choice in Switzerland. *disP-Plann. Rev.* (July 2015), 1–25.
- Aultman-Hall, L., Kaltenecker, M.G., 1999. Toronto bicycle commuter safety rates. *Accid. Anal. Prev.* 31 (6), 675–686.
- Bahamonde-Birke, F., Kunert, U., Link, H., Ortúzar, J.D.D., 2015. About attitudes and perceptions: finding the proper way to consider latent variables in discrete choice models. *Transportation* 42 (6), 1–19.
- Bartczak, A., Mariel, P., Chilton, S., Meyerhoff, J., 2016. The impact of latent risk preferences on valuing the preservation of threatened lynx populations in Poland. *Aust. J. Agric. Resource Econ.* 60 (2), 284–306.
- Ben-Akiva, M., Walker, J., Bernardino, A., Gopinath, D.A., Morikawa, T., Polydoropoulou, A., 2002. Intergration of choice and latent variable models. In: Mahmassani, H. (ed.), *In Perpetual Motion: Travel Behavior Research Opportunities and Application Challenges*. Pergamon, edition, vol. 1, pp. 431–470.
- Bernhof, I.M., Carstensen, G., 2008. Preferences and behaviour of pedestrians and cyclists by age and gender. *Transp. Res. Part F: Traffic Psychol. Behav.* 11 (2), 83–95.
- Bhat, C.R., 1997. An endogenous segmentation mode choice model with an application to intercity travel. *Transp. Sci.* 31 (1), 34–48.
- Bierlaire, M., 2003. BIOGEME: a free package for the estimation of discrete choice models. In: *Swiss Transport Research Conference, Ascona, Switzerland*, pp. 1–27.
- Broach, J., Dill, J., Gliebe, J., 2012. Where do cyclists ride? A route choice model developed with revealed preference GPS data. *Transp. Res. Part A: Policy Pract.* 46 (10), 1730–1740.
- Caulfield, B., Brick, E., McCarthy, O.T., 2012. Determining bicycle infrastructure preferences – a case study of Dublin. *Transp. Res. Part D: Transp. Environ.* 17 (5), 413–417.
- Chataway, E.S., Kaplan, S., Nielsen, T.A.S., Prato, C.G., 2014. Safety perceptions and reported behavior related to cycling in mixed traffic: a comparison between Brisbane and Copenhagen. *Transp. Res. Part F: Traffic Psychol. Behav.* 23, 32–43.
- ChoiceMetrics, 2012. *Ngene 1.1.2 User Manual & Reference Guide*.
- Chorus, C.G., Kroesen, M., 2014. On the (im-)possibility of deriving transport policy implications from hybrid choice models. *Transp. Policy* 36, 217–222.
- Daly, A., Hess, S., Patruni, B., Potoglou, D., Rohr, C., 2012. Using ordered attitudinal indicators in a latent variable choice model: a study of the impact of security on rail travel behaviour. *Transportation* 39 (2), 267–297.
- Dill, J., Gliebe, J., 2008. *Understanding and Measuring Bicycling Behavior: A Focus on Travel Time and Route choice*. Technical Report December, Oregon Transportation Research and Education Consortium.
- Dill, J., Goddard, T., Monsere, M. C., McNeil, N., 2015. Can protected bike lanes help close the gender gap in cycling? Lessons from five cities. In: *Transportation Research Board 94th Annual Meeting*.
- Emond, C.R., Tang, W., Handy, S.L., 2014. Explaining gender difference in bicycling behavior. *Transp. Res. Rec.* 2125, 16–24.
- Evans-Cowley, J.S., Akar, G., 2014. StreetSeen visual survey tool for determining factors that make a street attractive for bicycling. *Transp. Res. Rec.: J. Transp. Res. Board* 2468, 19–27.
- Fowkes, T., Wardman, M., 1988. The design of stated preference travel choice experiments: with special reference to interpersonal taste variations. *J. Transp. Econ. Policy* 22 (1), 27–44.
- González, F., Melo-Riquelme, C., de Grange, L., 2016. A combined destination and route choice model for a bicycle sharing system. *Transportation* 43 (3), 407–423.
- Grisolía, J.M., Willis, K.G., 2012. A latent class model of theatre demand. *J. Cult. Econ.* 36 (2), 113–139.
- Guevara, C.A., Polanco, D., 2016. Correcting for endogeneity due to omitted attributes in discrete-choice models: the multiple indicator solution. *Transportmetrica A: Transp. Sci.* 12 (5), 458–478.
- Heinen, E., van Wee, B., Maat, K., 2010. Commuting by bicycle: an overview of the literature. *Transp. Rev.* 30 (1), 59–96.
- Hess, S., Shires, J., Jopson, A., 2013. Accommodating underlying pro-environmental attitudes in a rail travel context: application of a latent variable latent class

- specification. *Transp. Res. Part D: Transp. Environ.* 25, 42–48.
- Hoffer, B., 2015. The Potential of Online Respondent Data for Choice Modeling in Transportation Research: Evidence from Stated Preference Experiments using Web-based Samples (Master thesis). ETH-Singapore Centre for Global Environmental Sustainability (SEC), Singapore.
- Hood, J., Sall, E., Charlton, B., 2011. A GPS-based bicycle route choice model for San Francisco, California. *Transp. Lett.: Int. J. Transp. Res.* 3 (1), 63–75.
- Horowitz, J.L., 1983. Statistical comparison of non-nested probabilistic discrete choice models. *Transp. Sci.* 17 (3), 319–350.
- Howard, C., Burns, E., 2001. Cycling to work in phoenix: route choice, travel behavior, and commuter characteristics. *Transp. Res. Rec.* 1773 (1), 39–46.
- Hoyos, D., Mariel, P., Hess, S., 2015. Incorporating environmental attitudes in discrete choice models: an exploration of the utility of the awareness of consequences scale. *Sci. Tot. Environ.* 505, 1100–1111.
- Hurtubia, R., Atasoy, B., Glerum, A., Curchod, A., Bierlaire, M., 2010. Considering latent attitudes in mode choice: the case of Switzerland. In: 12th WCTR, July 11–15, 2010-Lisbon, Portugal, pp. 1–17.
- Hurtubia, R., Guevara, C.A., Donoso, P., 2015. Using images to measure qualitative attributes of public spaces through SP surveys. *Transp. Res. Procedia* 11, 460–474.
- Hurtubia, R., Nguyen, M.H., Glerum, A., Bierlaire, M., 2014. Integrating psychometric indicators in latent class choice models. *Transp. Res. Part A: Policy Pract.* 64, 135–146.
- Kang, L., Fricker, J.D., 2013. Bicyclist commuters' choice of on-street versus off-street route segments. *Transportation* 40 (5), 887–902.
- Klobucar, M.S., Fricker, J.D., 2007. Network evaluation tool to improve real and perceived bicycle safety. *Transp. Res. Rec.* 2031 (-1), 25–33.
- Larsen, J., El-Geneidy, A., 2011. A travel behavior analysis of urban cycling facilities in Montreal, Canada. *Transp. Res. Part D: Transp. Environ.* 16 (2), 172–177.
- Lee, L.F., 1982. Specification error in multinomial logit models. Analysis of the omitted variable bias. *J. Econometrics* 20 (2), 197–209.
- Ma, L., Dill, J., 2017. Do people's perceptions of neighborhood Bikeability match reality? *J. Transp. Land Use* 10 (2), 1–18.
- Majumdar Bandhu, B., Mitra, S., Pareekh, P., 2015. Methodological framework to obtain key factors influencing choice of bicycle as a mode. *Transp. Res. Rec.: J. Transp. Res. Board* (2512), pp 110–121.
- McCullagh, P., 1980. Regression models for ordinal data. *J. Roy. Stat. Soc.* 42 (2), 109–142.
- McFadden, D., Train, K., 2000. Mixed MNL models for discrete response. *J. Appl. Econometrics* 15 (November 1998), 447–470.
- Menghini, G., Carrasco, N., Schüssler, N., Axhausen, K.W., 2010. Route choice of cyclists in Zurich. *Transp. Res. Part A: Policy Pract.* 44 (9), 754–765.
- Mertens, L., Van Dyck, D., Ghekiere, A., De Bourdeaudhuij, I., Deforche, B., Weghe, N.V.D., Van Cauwenberg, J., 2016. Which environmental factors most strongly influence a street's appeal for bicycle transport among adults? A conjoint study using manipulated photographs. *Int. J. Health Geographics* 15 (1), 31.
- Moritz, J., 1998. Adult bicyclists in the united states: characteristics and riding experience in 1996. *Transp. Res. Rec.* 1636 (98), 1–7.
- Motoaki, Y., Daziano, R.A., 2015. A hybrid-choice latent-class model for the analysis of the effects of weather on cycling demand. *Transp. Res. Part A: Policy Pract.* 75, 217–230.
- O'Neill, V., Hess, S., Campbell, D., 2014. A question of taste: recognising the role of latent preferences and attitudes in analysing food choices. *Food Qual. Prefer.* 32 (PC), 299–310.
- Ortúzar, J.D.D., Willumsen, L.G., 2011. *Modelling Transport*. fourth ed. John Wiley & Sons.
- Palma, D., Ortúzar, J.d.D., Rizzi, L.L., Guevara, C.A., Casaubon, G., Ma, H., 2016. Modelling choice when price is a cue for quality a case study with Chinese wine consumers. *J. Choice Modell.* 19, 24–39.
- Parkin, J., Wardman, M., Page, M., 2007. Models of perceived cycling risk and route acceptability. *Accid. Anal. Prev.* 39 (2), 364–371.
- Pickens, J., 2005. *Organizational Behavior in Health Care*. Jones and Bartlett Publishers, Sudbury.
- Potoglou, D., Palacios, J.F., Feijóo, C., 2015. An integrated latent variable and choice model to explore the role of privacy concern on stated behavioural intentions in e-commerce. *J. Choice Modell.* 17, 10–27.
- Reynolds, C.C.O., Harris, M.A., Teschke, K., Crompton, P.A., Winters, M., 2009. The impact of transportation infrastructure on bicycling injuries and crashes: a review of the literature. *Environ. Health* 8 (1), 47.
- Rossetti, T., 2017. *Modelación de preferencias por dise no de infraestructura ciclista utilizando variables latentes perceptuales*. Master's thesis, Pontificia Universidad Católica de Chile.
- Rossetti, T., Saud, V., Hurtubia, R., 2017. I want to ride it where I like: measuring design preferences in cycling infrastructure. *Transportation*. <http://dx.doi.org/10.1007/s11116-017-9830-y>.
- Ruud, P.A., 1983. Sufficient conditions for the consistency of maximum likelihood estimation despite misspecification of distribution in multinomial discrete choice models. *Econometrica* 51 (1), 225–228.
- Ryley, T., 2006. Estimating cycling demand for the journey to work or study in West Edinburgh, Scotland. *Transp. Res. Rec.-Ser.* (1982), 187–193.
- SECTRA, 2013. *Análisis del comportamiento de la demanda de infraestructura especializada para bicicletas*. Technical report.
- SECTRA, 2015. *Encuesta origen destino de viajes 2012*. Technical report, Ministerio de Transportes y Telecomunicaciones, Santiago, Chile.
- Sener, I.N., Eluru, N., Bhat, C.R., 2009. An analysis of bicycle route choice preferences in Texas, US. *Transportation* 36 (5), 511–539.
- Shankwiler, K.D., 2006. Development a framework for behavior assessment of bicycle commuters: a cyclist-centric approach (Ph.D. thesis). Georgia Institute of Technology.
- Shen, J., 2009. Latent class model or mixed logit model? A comparison by transport mode choice data. *Appl. Econ.* 41 (22), 2915–2924.
- Stinson, M.A., Bhat, C.R., 2003. Commuter bicyclist route choice: analysis using a stated preferences survey. *Transp. Res. Rec.* 1828, 107–115.
- Tilahun, N.Y., Levinson, D.M., Krizek, K., 2007. Trails, lanes, or traffic: valuing bicycle facilities with an adaptive stated preference survey. *Transp. Res. Part A: Policy Pract.* 41 (4), 287–301.
- Vega, R., Niehaus, M., Waintrub, N., Peña, C., Galilea, P., 2015. Cycle-path design standards: a qualitative characterisation approach for cyclist flows in Santiago. In: *Thredbo: International Conference Series on Competition and Ownership in Land Passenger Transport*.
- Vij, A., Walker, J., 2016. How, when and why integrated choice and latent variable models are latently useful. *Transp. Res. Part B: Methodol.* 90, 192–217.
- Wachtel, A., Lewiston, D., 1996. Risk factors for bicycle-motor vehicle collisions at intersections. *J. Saf. Res.* 27 (3), 195.
- Walker, J., Ben-Akiva, M., 2002. Generalized random utility model. *Math. Soc. Sci.* 43 (3), 303–343.
- Walker, J., Li, J., 2007. Latent lifestyle preferences and household location decisions. *J. Geogr. Syst.* 9, 77–101.
- Yáñez, F., Raveau, S., Ortúzar, J.d.D., 2010. Inclusion of latent variables in mixed logit models: modelling and forecasting. *Transp. Res. Part A: Policy Pract.* 44 (9), 744–753.