Incorporating the explicit role of psychological factors on mode choice: a hybrid mode choice model by using data from an innovative psychometric survey

Alejandro Tudela (atudela@udec.cl)
Department of Civil Engineering, Universidad de Concepción, Chile

Khandker M. Nurul Habib (khandker.nurulhabib@utoronto.ca)
Department of Civil Engineering, University of Toronto, Canada

Juan Antonio Carrasco (j.carrasco@udec.cl)
Department of Civil Engineering, Universidad de Concepción, Chile

Ahmed Osman (a.osman@utoronto.ca)
Department of Civil Engineering, University of Toronto, Canada

Abstract

Mode choice modelling has been shifting from a classical analysis based upon level of service, cost and socioeconomic explanatory variables, to one resting on individual factors which might also be affecting the decision process.

The aim of this work is to study the effect of incorporating explicitly psychological factors into the individual decision making process, estimating a hybrid discrete choice model. These psychological factors complemented the level of service and cost attributes.

Data set comes from an ad hoc survey designed to gather information regarding psychological aspects, such as attitudinal and affective factors, and habit. The survey was based on the framework proposed by Triandis. According to this theory, attitudinal, affective and social factors generate an intention, which is mediated by habit and contextual conditions to generate the observed conduct. The contextual conditions correspond to modes level of service and costs, socioeconomic characteristics of the individual, and trips constraints. Respondents corresponded to university staff, academic and clerical.

Modes and socioeconomic attributes, as well as psychological ones, through a latent variables approach, were considered in the estimations. In spite of the relative small sample size (231 records), results related to a full model are quite promising when compared with a simpler model. The consideration of the full set of information through the hybrid model allowed for the disentangling of effects, showing more evidently the role of modes and socioeconomics attributes and psychological variables on mode choice.

Keywords
Latent variable, hybrid model, psychological attribute, simulation, mode choice
1 Introduction

The aim of this work is to study the role of psychological factors into the individual decision making process, using explicit measures of them, which are fitted in a hybrid discrete choice model. These factors complement traditional level of service attributes, such as time and cost.

Data come from an ad hoc survey designed to gather psychological aspects, such as attitude, affective factor and habit. The survey was based on the framework proposed by Triandis (1977). According to this theory, attitudinal, affective and social factors generate an intention, which is mediated by habit and facilitating conditions to generate the observed conduct. Facilitating or contextual conditions are modes level of service and costs, socioeconomic characteristics of individuals, and trips constraints. Respondents corresponded to university staff, academic and clerical, with a final sample size of 231 records. Apart from the psychological attributes, modes attributes and socioeconomic information was also collected.

Modes and socioeconomic attributes, as well as psychological ones, through a latent variables approach, were considered in the estimations. In spite of the relative small sample size, results related to a full model are promising when compared with a simpler model. The consideration of the full set of information through the hybrid model allowed for the disentangling of effects, showing more evidently the role of modes and socioeconomics attributes and psychological variables on mode choice.

The paper is organized as follows. A review of the theoretical background regarding the incorporation of psychological attributes into the demand analysis is given in the next section. The mathematical model is explained in section 3, whilst a description of the data is provided in section 4. Results are provided in section 5, whereas the main conclusions and comments are reported in the last section.

2 Background

Since the 1980s, the literature has been suggesting that an explicit consideration of psychological factors might help us to understand the transport decision making process by individuals (Stringer, 1981), complementing the standard compensatory approach (Hensher et al., 2005).

The explicit consideration of psychological factors into the discrete choice modelling was facilitated in the 1990s with the incorporation of latent variables (Ben Akiva et al., 2002). Hybrid models, together with sophisticated mathematical formulations and computing processes, have been the methodological response to the simultaneous consideration of level of service, cost, socioeconomic, and latent attributes (Svenson, 1998; Johannson et al., 2006; Bolduc et al., 2008).

An alternative approach to the hybrid one is to consider these attributes into the choice analysis through Structural Equation Modelling (Bollen, 1989; Golob, 2003). Applications of SEM into travel behaviour considering psychosocial attributes can be found in Carrus et al. (2008), Sakano
and Benjamin (2008), Ory and Mokhtarian (2009), Farag and Lyons (2010), Habib et al. (2010), Kattan et al. (2010) and Galdames et al. (2011).

In spite of these advances, little accent has been put on measuring explicitly these psychological factors and improving our understanding about their relevance. In this context, experimental work has used aggregated measures of these factors when intending to capture their role. However, Social Psychology provides theoretical frameworks to take into account explicitly the effect of different psychological and contextual conditions upon the decision making process (Steg et al., 2001; Forward, 2004; Steg, 2005; Anable et al., 2006; Gardner and Abraham, 2008).

One of these social psychology frameworks corresponds to the Theory of the Interpersonal Behaviour (TIB), by Triandis (1977). This theory suggests that observed behaviour corresponds to an intention that is mediated by habit (frequency of use) and facilitating conditions (contextual situation). Facilitating conditions are mode availability, level of service and cost related attributes, individual socioeconomic context, journey restrictions, and so on, whilst intention depends on three factors: attitude, affect, and social; they might interact among them (see Figure 1). The impact of these factors and components upon behaviour or intention have been considered in several studies (Verplanken et al., 1997; Bämberg and Schmidt, 2003; Fujii and Gärling, 2003; Bogers et al., 2005; Cantillo et al., 2007; Ettema and Verschuren, 2007; Hensher and Pucket, 2007; Domarchi et al., 2008; Escobar and Tudela, 2009; Svennson, 2009; Galdames et al., 2011; Habib et al., 2011).

Attitude is related to the importance and value assigned to an object, for instance, the transport mode being used or that could be used. Expectancy-value theory is deemed to explain attitude, indicating that the force of an attitude depends on these two elements (Reeve, 2005; Gärling et al., 1998).

The affective factor is associated with the emotional value assigned to the object. Affective factors consider the induced emotions related to an action, such as travelling. An affective appraisal consists of attributing an affective quality to something, which can lead to carrying out certain attitude (Gärling et al., 1998; Anable and Gatersleben, 2005).

The social factor and personal trait can be represented through three components: social role (what should be done by someone given his/her situation in a group or society), social norm (how others’ behaviour affects someone behaviour), and self-concept (how someone sees him/herself). The role of these components on conduct has been studied by Anable (2005).

Habit is associated with a repetitive conduct, with suboptimal features, emerging from a stable context but being generalized to many other situations. This generalization implies that it could be difficult to change conduct just through persuasion (Bämberg, 2003; Gardner, 2009).

Whereas behaviour and contextual conditions can be observed, asked or modelled, factors influencing intention as well as habit can be measured using psychometric tools. Attitudinal and social factors can be measured using Likert scales (Cohen and Swerdlik, 2002) whilst affective factors can be determined indirectly using Osgood semantic differential (Osgood et al., 1975), which rests on catching the meaning of an object for an individual using a list of bipolar
adjectives (perfect antonyms), which are utilized to assess the object. Habit can be calculated using Verplanken’s response-frequency questionnaire, measuring habitual behaviour intensity by counting the number of times behaviour would be repeated under specific circumstances without too much reasoning (Verplanken et al., 1994).

This paper uses the TIB framework to estimate a hybrid model, using psychosocial and contextual information gathered using an *ad hoc* questionnaire.

![Diagram of Theory of the Interpersonal Behaviour](image)

**Figure 1:** Framework of the Theory of the Interpersonal Behaviour

### 3 Mathematical model

Psychological factors, habit and facilitating conditions can be integrated in a hybrid discrete choice model to study revealed choice mode responses. The consideration of psychological factors is done through latent variables. These variables can be expressed as a function of socioeconomic and personal attributes, which are captured using appropriate designed surveys, which would reflect the traveller’s psychology.

However, since the psychological states cannot be directly predicted and observed, it is considered that the latent variables have a distribution, rather than fixed values. Jointly modelling mode choice probability considering latent variables as explanatory variables leave the likelihood function of a Logit mode choice model as a non-closed form expression. Such type of choice model with random latent variables as independent variable in the utility function is called hybrid choice model. Simulation estimation is necessary in this case. In this paper, we develop such a hybrid choice model using psychometric information. The model is estimated by using simulated likelihood estimation technique.
Consider that the individual traveller’s utility \((U)\) of choosing the \(j\)-th mode is given by the following utility function:

\[
U_j = \beta_j x_j + (\eta + \varepsilon' \theta')A + (\nu + \varepsilon'' \theta'')B + \alpha_j y + \varepsilon_j \quad j = 1, \ldots, n
\]

Here, \(x\) indicates a vector of observed variables, \(A\) indicates expectation, \(B\) indicates habit and \(y\) indicates a latent affective factors. \(\beta\) and \(\alpha\) are utility coefficients associated with \(x\) and \(y\). With respect to expectation and habit, the related coefficients represent random coefficients across the population. In these random coefficients, \(\eta\) and \(\nu\) indicate mean values of the corresponding coefficients and \(\varepsilon' \theta'\) and \(\varepsilon'' \theta''\) represent variances \((\theta'\) and \(\theta'')\) multiplied by standard normal variables \((\varepsilon'\) and \(\varepsilon''\)). The error term, \(\varepsilon_j\), indicates a random error component to capture the unobserved and random component of the utility function of the corresponding alternative.

The following function allows for capturing the variability of latent affective factor across the population and, at the same time, quantifying it in terms of the observed variables:

\[
y = \gamma z + \theta
\]

Here, \(z\) indicates a vector of observed variables and \(\gamma\) indicates their corresponding coefficients. It is considered that the random component, \(\theta\), captures unobserved and random factors influencing mode choice. Here, latent variables are measured on a longitudinal scale considering several characteristics and can be expressed as:

\[
M_k = v_k (\gamma z + \theta) + \psi_k
\]

In this case, \(v_k\) indicates the coefficient of the latent perception corresponding to \(k\) criteria, and \(\psi_k\) is a normal error term with a zero mean and \(\tau_k\) variance. Since the latent variables are random in nature, incorporating them into the mode choice utility leaves the likelihood function as non-closed form. Similarly, the directly measured expectation and habit variable are also considered endogenous variables and further expressed as functions of observed variables:

\[
A = \gamma' z' + \pi'
\]
\[
B = \gamma'' z'' + \pi''
\]

Here \(z'\) and \(z''\) indicate vectors of observed variables; \(\gamma'\) and \(\gamma''\) indicates their corresponding coefficients. In this case the random components \(\pi'\) and \(\pi''\) are normal error terms with zero mean with \(\tau'\) and \(\tau''\) variances correspondingly.

Considering endogenous expectation and habit with random coefficients as well as latent affective factors together with measured indications \((M_k)\), the joint likelihood function of the integrated choice model no longer remain in closed form. However, a simulated likelihood technique is used in this case (Walker, 2001; Habib et al., 2010). The resulting simulated likelihood function stands as:
\[
L_i = \frac{1}{D} \sum_{d=1}^{D} \exp\left( \sum_{j=1}^{J} \left( \beta_i x_j + (\eta + \varepsilon_d' \theta'j) A + (\nu + \varepsilon_d'' \theta''j) B + (\gamma d + \theta_d') \cdot \alpha_j + \xi_{jd} \right) \right)
\]

\[
D \text{ indicates the total number of iterations used in simulation estimation, whilst subscript } d \text{ refers to individual iterations.}
\]

In this joint likelihood function, it is necessary to restrict certain parameters to ensure identification of the model as well as to reduce the number of parameters to be estimated by using a relatively small data set. The obvious approach, without sacrificing too much of the model’s explaining capacity, is to assume variances (\(\tau_k\), \(\tau'\) and \(\tau''\)) equal to one. The model is estimated using a GAUSS code written for simulated likelihood function and gradient search algorithm (BFGS). Halton sequence is used for random numbers. For simulation, 1000 iterations were used to ensure stable parameter estimates.

4 Data Description

Data used in this study come from two surveys designed and collected in 2007 (Domarchi, 2007) and 2008 (Escobar, 2008). Respondents were a random sample of University staff: lectures, researchers and clerical officers, which were contacted in their working place. They were asked to answer a revealed preference questionnaire regarding the trip they have done during the morning in their way from home to work, with an initial sample of 409 records. Modes’ level of service and cost attributes, and users’ socioeconomic and psychometric data were gathered using ad hoc questionnaires.

Attitude was measured using a combination of expectation and value, through 5-point Likert scales. The valuation of the affective factor was made using the Osgood semantic differential, consisting of the comparison of 16 pairs of antonym words. Habit was measured using the Verplanken’s response-frequency questionnaire, based upon 10 hypothetical situations.

Attitude was measured towards cars and buses, being or not being a user the day of the survey. The affective dimension was also recorded for the substitute mode in case the used more were not available.

After screening the 2007 and 2008 data bases, the final sample size reached the total of 231 records, with 190 for car users and 41 for public transport users. With respect to gender, 50.4% were men. The average age was 47.6 years old, with a standard deviation of 11 years. Regarding their occupation, 52% were academic staff and a 48% clerical staff. Some of the respondents have two available modes, whereas others have up to four: car driver, car passenger, bus and shared taxi.
5 Results

As stated before and explained in section 3, a hybrid choice model was estimated, integrating the level of service, cost, socioeconomic and psychological information to explain mode choice. The following table contains some two models. Model 1 is a simpler model and considers the mode attributes, socioeconomic information and some of the psychological data, whereas Model 2 considers all the available information.

It must be remembered that the sample size was 231 records, a relative small sample for such a complex model, having an incidence on t test. Nevertheless, some of them are above the significance level at 95% confidence interval. Besides, it should be noticed that the log-likelihood (LL) for Model 2 is the joint one, which includes 16 indicator functions of the hybrid choice model. If the marginal LL were required, a 16 time-integration over 16 indicator functions should be done.

When checking for the traditional variables: cost and level of service ones, it can be seen that sign are according to what it is expected. However, cost is one of the less explicative, when compared with the level of service ones. Focusing on the different times, it can be observed that the in-vehicle and waiting times are more relevant than walking time, something that might be due to the level of access and penetration related to the public transport network.

As usual, car ownership and income have a positive effect upon car use; once someone has got a car, he/she will use it.

When estimating the full model, it can be observed that the size of some coefficients changes, as well as their level of significance. In general, it can be said that the incorporation of the full information allows for the detection of the real role of some of the variables on the choice process. For instance, the cost parameter gets smaller and less significant, like the walking time one, whereas the in-travel and waiting times coefficients increase their magnitude and significance. A similar effect occurs with the car ownership and income variables. Their coefficients size gets reduced, but they are more significant.

Besides, it can be noticed that for those variables which have associated a distribution, the standard deviation is smaller for Model 2, with a lower significance. This would imply that the incorporation of the full information reduces the variation of the mean of those variables, id est, there would be more certainty with respect to the mean value for the variable. Randomness would be caught by the new variables added to the Model 2 when compared with Model 1.

With respect to the role of the psychological variables, it can be said that the expectancy – how good is the transport mode being used – resulted higher for car users than for public transport users, something that has been founded in previous studies in the same area. Regarding habit, this has a base component, being stronger for males, and increasing with age. Finally, the affective factor (emotional appraisal) would be lower for men, and decreasing with age. Due to variability and randomness in responses, the random error term for the affective factor resulted significant.
Table 1: Estimation results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1</th>
<th></th>
<th>Model 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mode</td>
<td>Parameter</td>
<td>t test</td>
<td>Parameter</td>
</tr>
<tr>
<td>Alternative Specific Constant</td>
<td>Car Driver</td>
<td>-2.190</td>
<td>-0.52</td>
<td>-3.230</td>
</tr>
<tr>
<td></td>
<td>Car Passenger</td>
<td>-0.971</td>
<td>-1.07</td>
<td>-6.846</td>
</tr>
<tr>
<td></td>
<td>Bus</td>
<td>-0.263</td>
<td>-0.13</td>
<td>1.536</td>
</tr>
<tr>
<td>Cost/Household Income</td>
<td>All Modes</td>
<td>-0.461</td>
<td>-0.88</td>
<td>-0.071</td>
</tr>
<tr>
<td>In-Vehicle Travel Time</td>
<td>All Modes</td>
<td>-0.063</td>
<td>-1.15</td>
<td>-0.132</td>
</tr>
<tr>
<td>Waiting Time</td>
<td>Bus and Shared Taxi</td>
<td>-0.167</td>
<td>-1.07</td>
<td>-0.490</td>
</tr>
<tr>
<td>Walking Time</td>
<td>Bus and Shared Taxi</td>
<td>0.141</td>
<td>1.20</td>
<td>-0.073</td>
</tr>
<tr>
<td>Total Cars in Household</td>
<td>Car Driver</td>
<td>1.371</td>
<td>1.40</td>
<td>0.958</td>
</tr>
<tr>
<td>Personal Income/Household Income</td>
<td>Car Driver</td>
<td>2.926</td>
<td>1.44</td>
<td>2.341</td>
</tr>
<tr>
<td>Latent Expectation: Car</td>
<td>Mean</td>
<td>Car Driver</td>
<td>-0.193</td>
<td>-0.27</td>
</tr>
<tr>
<td></td>
<td>Standard Deviation</td>
<td>Car Driver</td>
<td>0.333</td>
<td>1.03</td>
</tr>
<tr>
<td>Latent Expectation: Public Transport</td>
<td>Mean</td>
<td>Bus</td>
<td>-0.349</td>
<td>-0.75</td>
</tr>
<tr>
<td></td>
<td>Standard Deviation</td>
<td>Bus</td>
<td>0.251</td>
<td>1.29</td>
</tr>
<tr>
<td>Habit: Car</td>
<td>Mean</td>
<td>Car Driver</td>
<td>2.944</td>
<td>1.01</td>
</tr>
<tr>
<td></td>
<td>Standard Deviation</td>
<td>Car Driver</td>
<td>2.057</td>
<td>1.32</td>
</tr>
<tr>
<td>Affective Factor</td>
<td>Car Driver</td>
<td>-8.764</td>
<td>-2.86</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Car Passenger</td>
<td>-7.002</td>
<td>-2.94</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bus</td>
<td>0.116</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Shared Taxi</td>
<td>1 (Fixed)</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

Measurement Model

| Latent Expectation: Car | Constant | 4.802 | 4.26 |
| Gender: Male           | 0.083    | -0.62 |
| Logarithm of Age       | -0.088   | -0.30 |
| Latent Expectation: Bus| Constant | 3.319 | 2.94 |
| Gender: Male           | 0.102    | 0.77  |
| Logarithm of Age       | 0.151    | 0.51  |
| Latent Habit: Car      | Constant | 0.486 | 0.43 |
| Gender: Male           | 0.085    | 0.64  |
| Logarithm of Age       | 0.069    | 0.23  |
| Latent Affective Factor| Gender: Male | -0.049 | -0.77 |
| Logarithm of Age       | -0.229   | -10.98|
| Random Error           | 0.511    | 10.55 |

Indicator Function

| Good               | Latent Affective | 2.572 | 12.63 |
| Simple             | Latent Affective | 1.819 | 11.80 |
| Strong             | Latent Affective | 1.458 | 11.24 |
| Deserted           | Latent Affective | -0.339| -4.31 |
| Comfortable        | Latent Affective | 2.431 | 12.57 |
| Safe               | Latent Affective | 2.343 | 12.48 |
| Quick              | Latent Affective | 2.450 | 12.54 |
| Pleasant           | Latent Affective | 2.549 | 12.65 |
| Great              | Latent Affective | 1.092 | 10.08 |
| Clear              | Latent Affective | 1.625 | 11.55 |
| Active             | Latent Affective | 1.685 | 11.70 |
| Flexible           | Latent Affective | 1.574 | 11.48 |
| Clean              | Latent Affective | 1.954 | 12.07 |
| Quite              | Latent Affective | 1.861 | 11.98 |
| Big                | Latent Affective | 1 (Fixed) | - |
| Known              | Latent Affective | 2.273 | 12.44 |

LL of Joint Choice and Latent Variable Model | -7982|
LL of Choice Model Only | -88    | -99   |
LL of Null Model        | -219   | -219   |
Rho-Squared Value       | 0.60   | 0.55   |
6 Conclusions and comments

The estimation of a hybrid choice model, considering psychological attributes among other variables, measured using an ad hoc questionnaire and introduced into the choice model as latent variables, showed to be feasible in spite of the fairly small sample size. Attitudinal and affective factors, as well as habitual behaviour, were quantified and considered into the model.

The incorporation of the psychological variables into the modelling allowed for the unravelling of a different role for the level of service and cost attributes. When comparing two estimations, one considering the full information available with respect to a simpler one, it was found that the relative weight of the cost is reduced, whereas the relative importance of the in-vehicle and waiting times are increased. Car ownership and income level are still affecting car choice, but they are relatively less important than they are when compared with a simpler model.

The impact of the attitudinal and affective factors, as well as habit, on mode choice could be studied explicitly, avoiding the usage of other approaches to consider these factors, such as dummy variables, or structural equation modelling.

Further work is needed in terms of modelling, considering extra information available in the original data base, as well as introducing some interactions between socioeconomic and level of service attributes.

Certainly, a larger data base is required given the complexity of the model, incorporating the social factor indicated by Triandis framework into the data collection and analysis. Social norm, role and self concept indeed might affect some people’s behaviour in relation with mode choice decision making.

Finally, a complete model might allow us to analyse and assess the effectiveness of some transport policies, which are being introduced or studied in urban areas, such as road pricing schemes, bus priority investments, etc. It has been shown that time, cost and socioeconomic factors are not the only attributes affecting people’s decision making in the short run.

References


