Modelling travel time perception in transport mode choices

Silvia F. Varotto (corresponding author)
Dipartimento di Ingegneria e Architettura
Università degli Studi di Trieste
Via Valerio 6/4, 34127 Trieste, Italy
Currently:
Department of Transport and Planning
Faculty of Civil Engineering and Geosciences
Delft University of Technology
Stevinweg 1, P.O. Box 5048, 2600 GA Delft, The Netherlands
Phone: +31 (0)15 2789575, Fax: +31 (0)15 2783179, Email: s.f.varotto@tudelft.nl

Aurélie Glerum
Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne
Station 18, 1015 Lausanne, Switzerland
Phone: +41 (0)21 6932435, Fax: +41 (0)21 6938060, Email: aurelie.glerum@epfl.ch

Amanda Stathopoulos
Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne
Station 18, 1015 Lausanne, Switzerland
Currently:
McCormick School of Engineering & Applied Science
Northwestern University, TECH A335
2145 Sheridan Rd, Evanston, IL 60208
Phone: +1 847-491-5629, Email: a-stathopoulos@northwestern.edu

Michel Bierlaire
Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne
Station 18, 1015 Lausanne, Switzerland
Phone: +41 (0)21 6932437, Fax: +41 (0)21 6938060, Email: michel.bierlaire@epfl.ch

Giovanni Longo
Dipartimento di Ingegneria e Architettura
Università degli Studi di Trieste
Via Valerio 6/4, 34127 Trieste, Italy
Phone:+39 040 5583576, Fax: +39 040 5583580, Email: longo@dica.units.it

<table>
<thead>
<tr>
<th>Word count:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>243</td>
</tr>
<tr>
<td>Main text</td>
<td>5412</td>
</tr>
<tr>
<td>Figures</td>
<td>3 (750)</td>
</tr>
<tr>
<td>Tables</td>
<td>2 (500)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>6905</strong></td>
</tr>
<tr>
<td>References (n)</td>
<td>31</td>
</tr>
</tbody>
</table>

Paper submitted to TRB Annual Meeting 2015
November 14th, 2014
ABSTRACT

Travel behaviour models typically rely on data afflicted by errors, in perception (e.g., over/under-estimation by traveller) and measurement (e.g., software or researcher imputation error). Such errors are shown to have a relevant impact on model outputs. So far a comprehensive framework to deal with different types of biases related to travel model inputs is missing.

In this paper, focusing on travel time, we include two types of measures (i.e., "calculated" and "reported") as indicators of an unobservable true travel time. The aim of including these travel time indicators is to investigate how the underlying travel time perception on behalf of travellers influences the modal choice, compared to the role of externally obtained measurements. The model framework is a latent variable structure where the different travel time indicators serve as manifestations, characterized by different types of biases, of the true travel time. The model is applied to a mode choice case study from Trieste (Italy).

Notably, for this data-set, it is established that the calculated travel time distributions (i.e., measured by devices such as an assignment model developed with the software Visum and Google Maps) do not match the reported travel time distributions (i.e., reported by respondents in the survey). Therefore, a discrete choice model that employs available data and accounts for data limitations is developed. Results from comparing the base model assuming error-free inputs and the integrated models show more consistent and plausible model outputs such as value of time for the latter.

Keywords: travel time perception, measurement error, mode choice, discrete choice.
INTRODUCTION

It is well established in the travel behaviour field that travellers often overestimate or underestimate the true travel time of their trip. Such perception errors will then influence their travel decisions. Similarly when travel attributes, such as time, are imputed using softwares there may be relevant measurement errors. These measures will typically be used in travel choice models. In addition, the "calculated" travel time distributions (i.e., calculated by softwares) could differ from respondent-declared "reported" travel time distributions. While these issues have been raised in the literature there is still a dearth in understanding as to what constitutes the ideal framework to deal with the different types of biases related to each source of travel time model inputs. There is no clear view of the consequences of disregarding the measurement errors related to each data-source.

The subjective nature of temporal judgement is established in psychological research. Hornik (1) for instance finds that good mood leads to retrieving biased memories of time congruent with the mood. In transportation research, Bates et al. (2) argue that it is likely that travellers are maximizing utility according to their own divergent views of the travel time distribution notwithstanding actual measurements. Consequently, travellers will differ in their optimal choices depending on the degree of distortion of their subjective distribution with regards to the actual measurement distribution. In relation to this, Rietveld (3) notes that in travel surveys most respondents apply rounding of departure and arrival times to multiples of 5, 15 and 30 minutes. A possible explanation for this effect is that scheduled activities force people to plan their trips in advance which provide them with anchor points for their memory afterwards. Explicitly addressing rounding leads to a considerably better treatment of variances of reported travel times and enables one to avoid biases in the computation of average transport times based on travel surveys.

Although there is a broad literature on measurement errors in the econometric literature, few researches are directly addressing measurement errors in transportation modelling and in choice models. McFadden (4) notes that aggregate travel time data are often not sufficient and individual measurements of travel times are fundamental for modelling travel behaviour.

In the last decade, the popularity of hybrid choice models has grown considerably in a wide number of disciplines, including transport (5, 6, 7, 8, 9). Integrated Choice and Latent Variable models (ICLV) are primarily employed for including attitudes and perceptions as explanatory variables of the choice, using psychometric scales as indicators of unobservable latent constructs (10, 11, 12). This methodology could potentially be used to deal with any type of variable which affects the choice. Walker et al. (13) focus on how to estimate travel demand models when the underlying quality of level of service data (times and costs) are poor. It is demonstrated that a choice model with measurement errors results in inconsistent estimates of the parameters and therefore, methods to correct measurement errors need to be employed. The authors propose to use the hybrid choice framework to integrate travel time as a latent variable and use the measured travel time as an indicator of the latent true travel time. The ICLV model for true travel time leads to significant shifts in both the travel time and travel cost parameters, resulting eventually in a large increase in the value of time. In the context of their case study, the VOT calculated with the hybrid choice model seems to be more realistic than the VOT calculated with the base model. Hess et al. (14) develop a latent variable approach to deal with missing values and measurement errors relating to income. The reported income is replaced by a latent income variable in a choice model, using the stated income as an indicator of the unobservable true income in a measurement model. In contrast with using imputation of missing values, the simultaneous estimation with the choice model allows the observed choices to affect the latent variable. Furthermore, unlike approaches relying on stated income or on imputed values, the method is directly applicable for forecasting. Indeed, the approach of estimating separate cost sensitivities for respondents with missing income does not easily carry over into forecasting when the forecast population is different from the estimation sample.

Turning to examine missing observations, research has shown that imputing such values typically generates additional error. This applies both when missing data is treated analytically (15) or when applying multiple imputation (16) originally proposed by Rubin (17). Multiple imputations can be used when accurate data for a subsample of the observations are available. Bhat (18) imputes a continuous variable for missing values, meaning that the variable is drawn from the observed variables.
The aim of this paper is to propose a new model framework that relies on existing, and potentially biased data (subjective reported or instrument imputed). The scope is to develop demand models that overcome the inherent limitations in the data and can provide more robust estimates and enhanced forecasting accuracy compared to standard models that treat the imprecise values as true.

The issues of poorly measured levels of service data and not detailed enough travel behaviour information are explored in a real mode choice case study for a university campus in Trieste (Italy). It appears important to note that the above-mentioned dataset was collected for an assignment model whereas no discrete choice model has previously been estimated. Public transport network levels of service measures are estimated by means of an assignment model developed with the software Visum. The imputation of travel time using such a model could potentially lead to large measurement errors. In addition, the travel time distributions calculated with the help of instruments (i.e., “calculated travel time”) could differ from the travel time distributions reported by respondents (i.e., “reported travel time”). In order to deal with the above-mentioned limitations, the ICLV framework is proposed to correct measurement errors in a mode choice context (13, 14).

The paper is structured as follows. The next section presents the transport mode choice case study and the data processing procedure. The section after that provides the methodology and includes the model specification regarding the Multinomial Logit model and the Integrated Choice and Latent Variable models. The section which follows this presents the discussion of the estimation results obtained by using the extended software package BIOGEME (19). This section is followed by the validation and policy analysis. The last section provides the conclusions and suggestions for future research, while also discussing the limitations concerning the dataset used and possible extensions.

SURVEY AND DATA COLLECTION

A comprehensive data collection campaign was carried out in Trieste between November 2009 and January 2010, in the framework of UniMob - a Mobility Management project for the University of Trieste studying staff and students travel demand. The survey consisted of a quantitative part collecting socio-demographic data and information related to the role occupied, the length of service, the frequency of being at the university, their residence status and the means available. In addition, respondents were asked to thoroughly report information about the home-university trips made during the day, including their origins, destinations, chosen modes, arrival and departure times. Information on travel habits, activities during travel, potential reasons for mode switching, perception of the risks associated with each mode and opinions on urban mobility related topics were also gathered. The main data source used in this research consists of revealed preferences (RP) data on mode choice (20). The survey was performed as an on-line questionnaire, which allowed the whole university population (24685 users) to participate in the interview. During November 2009 all regularly registered students (21601 users) and all teaching and administrative staff (3084 users) were invited by email to complete the questionnaire (21). In total 3976 valid questionnaires were collected (response rate: 16.11%). Descriptive statistics are available in Varotto et al. (22).

Data processing

The available data have to be processed in order to extract all the variables necessary to define the utility functions for the alternative modes. The choice of the transportation mode is assumed to be among four alternatives: car, motorcycle, public transport (PT) and walk. All or some of these modes are available to each user, depending on their availability.

Origin-destination matrices (OD) are constructed using as reference a Visum assignment model (23). In this model, each zone was represented through a point placed in the barycentre of the zone and each trip was modelled as a travel between the barycentres of the corresponding OD zones. The OD couples are assigned to each user controlling for the residential status declared, the frequency with which each respondent went to the university, the modes used, the sequence of modes and the reported travel time.
Travel time is imputed for each alternative mode separately. Different devices are used such as the assignment model made by Visum and Google Maps. Distances between each origin and destination are calculated by using the website Google Maps and based on the addresses of origin and destination reported by respondents. Travel costs are calculated for the chosen and the unchosen alternatives. More details about the data processing are available in Varotto et al. (22). In addition, significant differences between the staff and student samples are discovered with formal comparisons and for the purpose of the current study only the more homogeneous sample of staff is selected.

Travel time indicators
As a first exploratory investigation of the underlying travel time perception, the travel durations reported by respondents for the chosen alternative are analysed and compared to the travel times obtained from the assignment software Visum (PT) and from Google Maps (car and motorcycle). For the chosen alternative two main indicators are used:

- Reported arrival and departure times;
- Calculated travel time (imputed using the above described procedures).

For the unchosen alternatives only the calculated travel time is available. Figure 1 illustrates the difference between the two measures for PT.

![Histogram of reported and calculated travel time for public transport.](image)

The comparison reveals that the means of reported and calculated travel time do not match for any modes. Specifically, the reported travel time is overestimated compared to the calculated for all modes except walk. Moreover, it can be observed in Figure 1 that the reported values are clustered around multiples of 5, 15, 30 and 60 minutes. This preliminary analysis points towards the relevance of developing methods to account for imprecision in travel time reporting.
**METHODOLOGY**

The aim of the present research is to investigate the impact of the underlying travel time perception of individual travellers on mode choices. Therefore, the unobservable true travel time which drives the choices is modelled as a latent variable. This approach aims to understand the consequences of the direct inclusion of assumed error-free time measures in mode choice modelling. In the following, two main models including different travel time indicators within the ICLV framework will be presented:

- Specification 1 using calculated time, assumed to be the best revealer of the true travel time and affected by different errors related to the instrument employed for imputation;
- Specification 2 using reported travel time, assumed instead to give better insight to the decision process and be affected by perceptual biases.

Results will be compared with a reference Multinomial Logit model (MNL) with an equivalent specification but using the calculated travel time directly in the utility function for all the modes. The methodology is employed to correct the travel time of public transport, because the imputation of travel time using the Visum assignment model could potentially lead to large measurement errors. In addition, the gap between the reported and calculated travel time seems to affect more the travellers who choose PT than all remaining alternatives.

**ICLV specification 1**

The first specification is an ICLV model assuming that the true travel time is a latent attribute and the calculated travel time can be used as an indicator. The model schematized in Figure 2 is called the ICLV for the true travel time, since a latent attribute (the true travel time) is integrated into the choice model. A similar framework was proposed by Walker et al. (13) to deal with measurement errors in the calculated travel time. Observable variables such as explanatory variables, indicators and choices are represented by rectangular boxes and latent variables such as utilities and latent attributes are represented by ovals. Structural equations are represented by straight arrows while measurement equations are represented by dashed arrows.

![Figure 2 ICLV-1 for true travel time for public transport.](Diagram)

- Distance;
- Travel cost;
- Travel time;
- Parking time in *Cattinara* and *Stazione*.

**Attributes of the alternatives**

- Female;
- Residents;
- Year of service;
- Indirect trip;
- Trips to other faculties;
- Faculty in *Città Vecchia*.

**Characteristics of the traveller**

- Choice: Car, Moto, PT, Walk.
- Indicator of Travel Time
- Calculated travel time

**Choice Model**

**Latent attribute model**

26 FIGURE 2 ICLV-1 for true travel time for public transport.
Latent Variable model: structural equation for latent attribute

In the Latent Variable model, the true travel time $TT_n^*$ (i.e., latent attribute) is assumed to be given by the equation (1) for the public transport alternative and for each individual $n$:

$$TT_n^* = c + \sigma \cdot \delta_n, \text{ with } \delta_n \sim N(0,1)$$

(1)

Where $c$ and $\sigma$ are parameter to be estimated. The mean of the true travel time $TT_n^*$ is represented by the parameter $c$. The error term has a mean equal to zero and a standard deviation equal to the parameter $\sigma$. The distribution of the latent variable $TT^*$ is $f_1(TT_n^*; c, \sigma)$.

Latent Variable model: measurement equation for latent attribute

Measurement equations are built with the corresponding indicators of travel time as given in the equation (2). $I_{1n}$ represents the indicator of calculated travel time for the respondent $n$:

$$I_{1n} = \alpha_1 + \lambda_1 \cdot TT_n^* + \sigma_1 \cdot \delta_{1n}, \text{ with } \delta_{1n} \sim N(0, \sigma_1)$$

(2)

Where $\alpha_1$ and $\lambda_1$ are parameters which are fixed for normalization purposes, $\sigma_1$ is a parameter to be estimated and $TT_n^*$ is the latent attribute. The error term has a mean equal to zero and a standard deviation equal to the parameter $\sigma_1$. The measurement equation is based on a continuous scale since the calculated travel time is continuous. The distribution of the indicator $I_1$ is $f_2(I_{1n} | TT_n^*; \alpha, \lambda, \sigma)$.

Discrete choice model

The latent attribute $TT_n^*$ is introduced into the utility function of the public transport alternative in place of the calculated travel time. It is essential to note that including the calculated travel time directly into the utility function assumes that the value is measured without error, while including the latent attribute $TT_n^*$ accounts for the distribution of the parameter. The utility $U_{in}$ of an alternative $i$ for a decision-maker $n$ is expressed as a function $V$ of observed characteristics $X_{in}, X_n$ and of the latent attribute $TT_n^*$ as given in the equation (3):

$$U_{in} = V(X_{in}, X_n, TT_n^*; \beta) + \varepsilon_{in}, \text{ with } \varepsilon_{in} \sim EV(0,1)$$

(3)

Where

- $X_{in}$ is a vector representing the attributes of the alternative $i$;
- $X_n$ is a vector representing the characteristics of the decision-maker;
- $TT_n^*$ is the unobservable true travel time;
- $\beta$ is a vector of parameter to estimate;
- $\varepsilon_{in}$ is the error term.

Integrated model framework

Under the assumption that the error terms are independent and $y_{in}$ is the indicator of choice for an individual $n$ (equal to 1 if alternative $i$ is chosen, and 0 otherwise), the likelihood function for an individual $n$ is given by the formula (4):

$$L_n(y_{in} | I_{1n}, X_{in}, X_n; \alpha, \lambda, \beta, \sigma) = \int_{TT^*} P(y_{in} | X_{in}, X_n, TT_n^*; \beta) \cdot f_2(I_{1n} | TT_n^*; \alpha, \lambda, \sigma) \cdot f_1(TT_n^*; c, \sigma) dTT^*$$

(4)

Where

- $P(y_{in} | X_{in}, X_n, TT_n^*; \beta)$ is the probability that alternative $i$ is chosen by individual $n$;
- $f_2(I_{1n} | TT_n^*; \alpha, \lambda, \sigma)$ is the density function of the indicator $I_1$;
- $f_1(TT_n^*; c, \sigma)$ is the density function of the latent attribute $TT_n^*$.
The parameters of the integrated model are estimated using maximum likelihood techniques as presented in equation (5):

$$\max_{\alpha, \lambda, \beta, \sigma} \sum_n \log \left( L_n \left( y_{in}, I_1 | X_{in}, X_n; \alpha, \lambda, \beta, \sigma \right) \right)$$  \hspace{1cm} (5)

ICLV specification 2

The second specification employs the reported travel time as an indicator of the true travel time. Aside from this, an important difference from the previous model is the inclusion of elements of travel behaviour in the measurement equation, assuming that they modulate the reported travel time explaining the perceived duration of the trip. For this motivation this inclusion is very central to the scope of more accurate models of travel time that effectively drives the choices. The specification is shown in Figure 3.

FIGURE 3 ICLV-2 for true travel time for public transport.

Latent Variable model: structural equation for latent attribute

In the Latent Variable model, the true travel time $TT_n^*$ (i.e., latent attribute) is assumed to be given by the equation (6) for the public transport alternative and each individual $n$:

$$TT_n^* = c + \lambda \cdot TT_n + \sigma \cdot \delta_n, \text{ with } \delta_n \sim N(0,1)$$  \hspace{1cm} (6)

Where $TT_n$ is the calculated travel time, $c$, $\lambda$ and $\sigma$ are parameter to be estimated. The distribution of the latent variable $TT_n^*$ is $f_1 (TT_n^*|TT_n; c, \lambda, \sigma)$. 

Latent attribute model

Choice Model

- Distance;
- Travel cost;
- Travel time;
- Parking time in Cattinara and Stazione.

Attributes of the alternatives

- Female;
- Residents;
- Year of service;
- Indirect trip;
- Trips to other faculties;
- Faculty in Città Vecchia.

Characteristics of the traveller

Choice: Car, Moto, PT, Walk.

Utility

True travel time

Indicator of Travel Time

Reported travel time

Calculated travel time

Elements of travel behaviour

- Music;
- Reading.
Latent Variable model: measurement equation for latent attribute

Assuming that elements of travel behaviour $X_n$ affect the reported travel time, the measurement equation for the indicator of reported travel time $I_{2n}$ for the respondent $n$ (with departure/arrival time) is built as given in equation (7):

$$ I_{2n} = \alpha_2 + \lambda_2 \cdot T\!T\!T_n^* + \beta_2 \cdot X_n + \sigma_2 \cdot \delta_{2n}, \quad \text{with } \delta_{2n} \sim N(0, \sigma_2) $$ (7)

Where $\alpha_2$ and $\lambda_2$ are parameters which are fixed for normalization purposes, $\beta_2$ and $\sigma_2$ are parameters to be estimated, $T\!T\!T_n^*$ is the latent attribute and $X_n$ are the socio economic variables and elements of travel behaviour of respondent $n$. The measurement equation is based on a continuous scale. The distribution of the indicator $I_{2n}$ is $f_2 (I_{2n} | X_n, T\!T\!T_n^*; \alpha, \lambda, \beta, \sigma)$.

For the construction of the measurement equation for the true travel time, which is assumed to be affected by elements of travel behaviour, a principal component analysis (PCA) is performed with the whole set of elements of travel behaviour as an exploratory step. The PCA is performed employing the statistical software R, using the package *psych* version 1.2.8 developed by Revelle (24). As a result two main travel behaviour features are included, as follows:

- **Reading\textsubscript{STAFF}**, corresponding to the statement “I never read during the home-university trip” (five-point Likert scale);
- **Music\textsubscript{STAFF}**, corresponding to the statement “I never listen to music during the home-university trip” (five-point Likert scale).

**ESTIMATION RESULTS**

The maximum likelihood method is used for model estimation, which is done by using the extended software package BIOGEME (19). The ICLV models are estimated using the sequential approach as described by Walker (8). The log-likelihood, the goodness of fit results and the estimation results are reported in Table 1. The final log-likelihood values and rho-bar-squared are calculated for only the choice component of the ICLVs to be comparable with the base model. The fit and the likelihood function of the choice component of the ICLVs are lower than in the MNL model. These results are consistent with recent findings by Vij and Walker (25). Indeed, they observed that any ICLV model can be reduced to a choice model without latent variables that fits the data at least as well as the original ICLV model from which it was obtained.

Looking at the utility parameters of the MNL model, explanations are provided discussing the signs and the magnitudes of the parameters related to the attributes and the other explanatory variables. All the parameters are statistically significant.

Referring to the parameters which regard the modal attributes of travel time, cost, distance and parking time, it can be observed that they affect the utility negatively, in line with expectations. For PT, the calculated travel time was missing for 14.21% users in this sample. Thus, we introduce a variable **MissingTime\textsubscript{PT,STAFF}** which is equal to 1 when the PT travel time is missing and 0 otherwise. For car alternative, a variation in travel time has a different effect on mode choice, depending on whether the distance travelled is short or long. For this motivation, the interaction $TT\!C\!A\!R/D$ is introduced into the utility function, implying that the coefficient of $TT\!C\!A\!R$ (i.e., $\beta_{TIME\_CAR\_STAFF}/D$) is scaled by distance $D$ expressed in kilometres in Table 1. This interaction is significant and improves the fitting for car alternative only. In addition, the parking time parameter is only significant for the staff members working at the faculties located near Ospedale Maggiore and near the railway station, where the parking lots available are limited and the time necessary to find a parking lot is high (e.g., 10 - 15 minutes).
Moreover, some socio-economic variables have a significant effect on the choice of transport modes. The parameters of the explanatory variables introduced have the expected signs and further observations are presented below:

- Female staff members tend to least prefer motorcycles among all modes; the negative sign of $\beta_{GENDER\_MOTO\_STAFF}$ implies that female staff is less likely to choose motorcycles than other modes;
- Staff members who reported more than 20 years of service are more likely to choose car than any other mode and this effect increases linearly with the number of years of service (positive $\beta_{20\_YEAR\_CAR\_STAFF}$). Possible explanations could be: first that older people prefer more to travel by car, second that people with a longer working experience could have a higher income and can afford a car;

### TABLE 1 Statistics and estimation results for the ICLV models

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Base model</th>
<th>ICLV specification 1</th>
<th>ICLV specification 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Statistics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of parameters</td>
<td>15</td>
<td>18</td>
<td>21</td>
</tr>
<tr>
<td>Number of parameters (choice component)</td>
<td>15</td>
<td>15</td>
<td>14</td>
</tr>
<tr>
<td>Number of observations</td>
<td>901</td>
<td>901</td>
<td>901</td>
</tr>
<tr>
<td>Final log-likelihood</td>
<td>-417.961</td>
<td>-490.554</td>
<td>-438.501</td>
</tr>
<tr>
<td>Adjusted rho-bar-squared</td>
<td>0.564</td>
<td>0.491</td>
<td>0.544</td>
</tr>
<tr>
<td><strong>Parameters</strong></td>
<td>Estimate</td>
<td>T-test</td>
<td>Estimate</td>
</tr>
<tr>
<td>$ASC_CAR_STAFF$</td>
<td>1.129</td>
<td>2.70</td>
<td>2.284</td>
</tr>
<tr>
<td>$ASC_MOTO_STAFF$</td>
<td>0.245</td>
<td>0.49</td>
<td>*</td>
</tr>
<tr>
<td>$ASC_PT_STAFF$</td>
<td>3.277</td>
<td>5.71</td>
<td></td>
</tr>
<tr>
<td>$ASC_SM_STAFF$</td>
<td></td>
<td>-</td>
<td>-0.170</td>
</tr>
<tr>
<td>$\beta_{COST_STAFF}$</td>
<td>-4.826</td>
<td>-11.98</td>
<td>-4.690</td>
</tr>
<tr>
<td>$\beta_{TIME_CAR_STAFF}$</td>
<td>-0.893</td>
<td>-5.59</td>
<td>-1.160</td>
</tr>
<tr>
<td>$\beta_{TIME_MOTO_STAFF}$</td>
<td>-0.142</td>
<td>-3.29</td>
<td>-0.110</td>
</tr>
<tr>
<td>$\beta_{TIME_PT_STAFF}$</td>
<td>-0.109</td>
<td>-7.13</td>
<td>-0.347</td>
</tr>
<tr>
<td>$\beta_{TIME_WALK_STAFF}$</td>
<td>-0.083</td>
<td>-9.33</td>
<td>-0.082</td>
</tr>
<tr>
<td>$\beta_{MISSING_TIME_PT_STAFF}$</td>
<td>-5.430</td>
<td>-6.62</td>
<td>-</td>
</tr>
<tr>
<td>$\beta_{MISSING_CAR_STAFF}$</td>
<td>-0.300</td>
<td>-4.85</td>
<td>-0.291</td>
</tr>
<tr>
<td>$\beta_{FEMALE_MOTO_STAFF}$</td>
<td>-0.956</td>
<td>-2.64</td>
<td>-0.974</td>
</tr>
<tr>
<td>$\beta_{20_YEAR_CAR_STAFF}$</td>
<td>0.018</td>
<td>2.35</td>
<td>0.018</td>
</tr>
<tr>
<td>$\beta_{INDIRECT_TRIP_CAR_STAFF}$</td>
<td>1.487</td>
<td>5.16</td>
<td>1.630</td>
</tr>
<tr>
<td>$\beta_{FACULTIES_PT_STAFF}$</td>
<td>0.787</td>
<td>3.03</td>
<td>0.616</td>
</tr>
<tr>
<td>$\beta_{CITTAVECCHIA_WALK_STAFF}$</td>
<td>0.780</td>
<td>1.96</td>
<td>0.664</td>
</tr>
<tr>
<td>$Latent\ time - \zeta$</td>
<td>-</td>
<td>-</td>
<td>19.466</td>
</tr>
<tr>
<td>$Latent\ time - \lambda$</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$Latent\ time - \sigma$</td>
<td>-</td>
<td>-</td>
<td>-0.551</td>
</tr>
<tr>
<td>$\beta_{READING_STAFF}$</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\beta_{MUSIC_STAFF}$</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\beta_{MISSING_READING_MUSIC}$</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Measurement Equation – $\sigma_1$</td>
<td>-</td>
<td>-</td>
<td>2.425</td>
</tr>
</tbody>
</table>
- Staff members who reported to have stopped once or more for different purposes during their home university trip are more likely to choose car. Possible explanations could be that first the necessity to stop for other purposes bring the need for more flexible forms of transport, second that the user may need to drive other passengers to different destinations.
- Staff members are more likely to choose PT than all the other modes in their home-university trip when they need to visit other faculties during the day. A possible explanation could be the difficulty to find an empty parking lot near the other faculties that should be visited during the day, than in the morning when the respondent arrives to university.

The estimated parameters of the MNL model and the choice component of the ICLV models are close to each other, with the exception of the time parameters. Walker et al. (13) note significant shifts in the time parameter of the ICLV model. Observing the time parameter of PT, which is associated to the latent time estimated, the magnitude increases in both ICLVs compared to the MNL.

Looking at the Latent Variable models, the means of the latent travel times are expected to have a magnitude and sign comparable to the mean values of the reported travel time (31.170 minutes) and of the calculated travel time (19.160 minutes) present in the raw dataset.

Referring to the ICLV-1, the mean of the latent time $c$ is equal to 19.466 minutes and the standard deviation $\sigma$ is 0.551 minutes. The measured travel time is assumed to be normally distributed with mean equal to the latent travel time and standard deviation $\sigma_1$ equal to 2.425 minutes. This standard deviation is significant, indicating that there is a measurement error inherent in the network derived travel times.

Referring to the ICLV-2, the mean of the latent time $c$ is equal to 22.445 minutes, the calculated travel time affects the latent time significantly with a magnitude $\lambda$ equal to 0.587 and the standard deviation $\sigma$ is 1.322 minutes. The reported travel time is assumed to be normally distributed with mean equal to the latent travel time and standard deviation $\sigma_1$ equal to 2.390 minutes. Moreover, the parameters $\beta_{READING\_STAFF}$ and $\beta_{MUSIC\_STAFF}$ referring to the elements of travel behaviour introduced in the measurement equation of the reported travel time have the expected signs, accordingly to the results obtained by the exploratory PCA. Clear conclusions for missing values cannot be drawn because the coefficient $\beta_{MISSING\_READING\_MUSIC}$ has a low t-statistic. The habit of never listening to music during the trip affects positively the reported time through the measurement equation, meaning that people who usually listen to music during the trip are more likely to report a travel time that is closer to the calculated one. The habit of never reading for leisure during the trip affects negatively the measurement equation of reported travel time, meaning that people who usually read for leisure during the trip are more likely to report a travel time that is far from the calculated one.

In general, it could be noted that the reported travel time seems to be an overestimation of the calculated travel time, in accordance with the results obtained through statistics.

**VALIDATION AND POLICY ANALYSIS**

First, a validation analysis of the model is provided in order to assess if the model could be applied to other potential data sets. Second, the value of time, which is an indicator of the willingness to pay (WTP) of individuals to reduce the duration of their trip by one hour, is computed for each alternative mode.

**Validation**

A proper validation of the model would require its application on a different data set but no other similar dataset is available. Therefore, the dataset available is split into two parts. First, 70% of the observations are selected randomly and the model is estimated on the latter. The ICLV models require a correction of the constants to reproduce correctly the choice probabilities within that sample (26). Second, the models are applied on the remaining 30% of the observations. The average number of alternatives available for each respondent and the corresponding chance level are calculated. Table 2 reports for each model the percentages of choice probabilities higher than 0.33 (chance level), 0.50, 0.70 and 0.90. The choice probabilities are well predicted by all three models. Cross-validation favours the MNL and ICLV-2 models for the “chance” model, while the ICLV-1 predicts choices better at higher levels of accuracy (where one alternative is utility-dominant).
TABLE 2 Percentages of choice probabilities higher than given thresholds and value of time

<table>
<thead>
<tr>
<th>Threshold</th>
<th>Base model</th>
<th>ICLV specification 1</th>
<th>ICLV specification 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>33 %</td>
<td>83.08 %</td>
<td>80.07 %</td>
<td>83.08 %</td>
</tr>
<tr>
<td>50 %</td>
<td>75.56 %</td>
<td>72.55 %</td>
<td>72.93 %</td>
</tr>
<tr>
<td>70 %</td>
<td>60.90 %</td>
<td>61.27 %</td>
<td>59.77 %</td>
</tr>
<tr>
<td>90 %</td>
<td>40.60 %</td>
<td>43.98 %</td>
<td>37.59 %</td>
</tr>
</tbody>
</table>

Value of Time

<table>
<thead>
<tr>
<th></th>
<th>Base model</th>
<th>ICLV specification 1</th>
<th>ICLV specification 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Car</td>
<td>4.69 €/h</td>
<td>6.27 €/h</td>
<td>4.94 €/h</td>
</tr>
<tr>
<td>Moto</td>
<td>1.76 €/h</td>
<td>1.40 €/h</td>
<td>1.62 €/h</td>
</tr>
<tr>
<td>PT</td>
<td>1.36 €/h</td>
<td>4.43 €/h</td>
<td>2.25 €/h</td>
</tr>
<tr>
<td>Parking search time</td>
<td>3.73 €/h</td>
<td>3.71 €/h</td>
<td>3.19 €/h</td>
</tr>
</tbody>
</table>

Value of time

The value of time (VOT) is presented in Table 2. Referring to the base model, it is important to point out that the VOT is higher for car than for all the other modes. The magnitude could be explained by the very low travel costs for private motorized modes (i.e., cost of fuel) and for public transport (i.e., 1-hour urban bus ticket equal to 1.10 €). In addition, it could be noted that the travel time considered only corresponds to the in-vehicle time for car, motorcycle and PT.

The VOT calculated for car seems consistent with the referential values for the VOT in Italy found in the literature. Indeed, many authors report a value of time for urban commuting trips by car around 4.00–5.00 €/h for users who work (27, 28, 29, 30). In addition, Rotaris et al. (31) estimate the VOT for students enrolled at the University of Trieste, developing a methodology which combines revealed and stated preferences: the VOTs vary from 1.40 to 2.80 €/h. The VOT obtained for PT seems to be lower than expected.

Referring to the ICLV models, the value of time is calculated for each mode because the correction for measurement errors implemented for public transport affects the whole model. It could be noted that, in both ICLVs estimated, the time parameter for PT shifts and the estimated VOT increases significantly, exactly as in the case study presented by Walker et al. (13). In the ICLV-1, the value of time increases by over 200%, from 1.36 €/h to 4.43 €/h. In the ICLV-2, the value of time increases by over 65%, from 1.36 €/h to 2.25 €/h. In addition, the 90% confidence bounds for the VOT of PT are calculated: in the base model, the confidence interval is equal to (1.15, 1.55); in ICLV-1 it is (3.57, 5.25); in ICLV-2 it is (1.58; 2.86). The confidence intervals do not overlap between the three models, meaning that different treatments of the time factor can lead to significantly different VOTs.

The VOTs obtained for PT within the hybrid choice framework seem to be more consistent with the referential values found in literature, compared to the values obtained for PT with the MNL model. Cherchi (28) points out that using the parameters estimated with a MNL, the VOT are largely underestimated and more realistic results could be obtained accounting for the variations in sensitivity among respondents. Fiorello and Pasti (30) report a value of time for PT trips performed by working commuters around 3.00 – 4.00 €/h. This value seems to be consistent with the VOT obtained in the ICLV-1. Cherchi (28) reports a VOT for urban trips performed by working commuters by PT around 2.00 €/h. A value equal to 2.00 €/h seems to be consistent with the VOT obtained in the ICLV-2.
CONCLUSION AND FUTURE RESEARCH

The aim of the present research is to develop methods for discrete choice modelling that account for limitations in available data. In order to deal with measurement errors in travel time, the use of the hybrid choice framework proposed by Walker (8) and Walker et al. (13) is explored: travel time is integrated into the choice model as a latent variable. This approach is applied on a data set from a university survey which was collected in Trieste (Italy) for an assignment model developed by Visum for public transport (23).

Two Latent Variable models for the value of travel time driving mode choices (i.e., “true travel time”) are integrated into the discrete choice model. The methodology is employed to correct the travel time of public transport, because of the accuracy of network derived level of service (i.e., travel time calculated by the assignment model in Visum) is expected to be lower for this alternative than all the other modes. In addition, the gap between the travel time which is reported by respondents (i.e., “reported travel time”) and the “calculated travel time” seems to affect more the travellers who choose PT. In the first Latent Variable model, the calculated travel time is assumed to be the best revealer and therefore used as an indicator of the true travel time. In the second Latent Variable model, the reported travel time is assumed to give better insight to the decision process and thus used as an indicator of the true travel time. In addition the reported time is assumed to be affected by elements of travel behaviour which influence the perceived duration of the trip, such as the habits of listening to music and reading during the home-university trip. The estimation results obtained by the MNL and by the ICLVs are compared based on statistical significance.

Referring to the two ICLVs estimated in the case study analysed, the second specification proposed, where the reported travel time is used as an indicator, fits the data better than the first one, where the calculated travel time is used as an indicator. However, the first specification has a slightly better forecasting power.

In terms of the analysis of demand, results indicate that the MNL which does not correct for measurement errors seem to underestimate travellers’ value of time (1.36 €/h). In addition, the value of time computed using the ICLV-1 (4.43 €/h) is higher than the value of time computed using the ICLV-2 (2.25 €/h). In comparison to the MNL, the ICLVs appear to produce more consistent parameters for the travel time variable which define more realistic travel demand indicators, closer to the referential values found in the literature for urban public transport in Italy, equal to 2.00 €/h (28) and 3.00-4.00 €/h (30).

The key point in this research is that measurement error can cause serious biases and methods that explicitly recognize and correct for such errors are necessary to improve the realism of the resulting analysis. There are many directions for future research: first, the ICLVs should be applied (e.g., calculating market shares and demand elasticities, and testing scenarios in which the calculated travel time by the assignment model is corrected in the LV models) and the implications on the resulting policy analysis should be investigated; second, a mixed discrete-continuous distribution of travel time could be introduced in the measurement equation of the Latent Variable model, explicitly addressing the rounding of reported travel time; third, the methodology could be used to correct for measurement errors for each mode; fourth, this framework has the potentiality to explicitly correct biases related to travel time reporting for the chosen and the unchosen alternatives (if available) using two different measurement equations.

In addition, the outcomes obtained suggest that a new survey should be carried out in order to collect more detailed information: first, further data should be collected to analyse more realistically the choice set of each user; second, more detailed disaggregate data regarding the level of service of public transport should be derived (access time, waiting time, number of transfers, in-vehicle time and the egress time); third, the respondent should be asked to report the travel time for both the chosen and the unchosen alternatives, specifying the access time, the in-vehicle time and the egress time; fourth, more detailed information regarding the household structure (such as income, number of people in the household, availability of private parking lot) seem to be necessary to better understand travel behaviour and to improve the predictive power of the choice models.

ACKNOWLEDGMENTS

The authors would like to thank Evanthia Kazagli at Ecole Polytechnique Fédérale de Lausanne for assistance and advice during the research.
REFERENCES

4. McFadden, D. Disaggregate Behavioral Travel Demand’s RUM Side: A 30-Year Retrospective. Presented at the International Association of Travel Behavior Research, Gold Coast, Queensland, Australia, 2000.


