Modelling travel time perception in transport mode choices

Silvia Francesca Varotto, Università degli Studi di Trieste
Aurélie Glerum, Ecole Polytechnique Fédérale de Lausanne
Amanda Stathopoulos, Ecole Polytechnique Fédérale de Lausanne
Michel Bierlaire, Ecole Polytechnique Fédérale de Lausanne

Conference paper STRC 2014

STRC | **14th Swiss Transport Research Conference**
Monte Verità / Ascona, May 14-16, 2014

Modelling travel time perception in transport mode choices

Silvia Francesca Varotto

Currently:
Department of Transport and Planning
Faculty of Civil Engineering and Geosciences
Delft University of Technology
Delft
The Netherlands

Phone: +31 (0)15 2789575
Fax: +31 (0)15 2783179
email: s.f.varotto@tudelft.nl

Research carried out at:
Dipartimento di Ingegneria e Architettura
Università degli Studi di Trieste
Trieste
Italia

Aurélie Glerum

Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne
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
Lausanne
Switzerland

Phone: +41 (0)21 6932435
Fax: +41 (0)21 6938060
email: amanda.stathopoulos@epfl.ch

Michel Bierlaire

Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne
Lausanne
Switzerland

Phone: +41 (0)21 6932537
Fax: +41 (0)21 6938060
email: michel.bierlaire@epfl.ch

April 2014

Abstract

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 software we may have relevant measurement errors. These measures will typically be used in travel choice models. In addition, the "calculated" travel time distributions (i.e. travel time calculated by instruments) 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, similarly, not a clear view of the consequences of disregarding the measurement errors related to each data-source.

In this paper we include both types of measurements (i.e. "calculated" and "reported") as indicators of an unobservable true travel time. The aim of including these various 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 distribution (i.e. travel time reported by respondents in the survey).

The main contribution of this work is the joint integration of travel time measurements derived from software or based on respondent perceptions in a latent variable structure. In addition, for respondent reported travel time indicators we test if these are affected by trip behaviour variables (e.g. habits of reading and listening to music during the trip). Current research efforts are dedicated to comparing models with different treatments of travel time components. The scope is to rigorously compare the impact in terms of derived elasticities, parameter ratios and forecasting performance.

Keywords

Discrete Choice Models – Transport Mode Choice – Measurement Errors – Latent Variables – Travel Time Perception

1. Introduction

Detailed land use and travel behaviour information are crucial in mode choice modelling but high quality data are not collected routinely in most cities. Nevertheless, standard approaches account only for observable variables to explain mode choices, such as the attributes of the alternatives and the socio-economic characteristics of the decision maker, ignoring measurement errors and correlations with unobservable factors. In addition, observations with missing values are removed or the missing information is imputed using as reference the existing complete observations.

Travel time could be considered one of the essential variables which affect modal choice and imprecision in the measurement of travel time can have a significant impact on estimates of travel demand indicators such as the value of time. Moreover, travellers usually underestimate or overestimate the actual travel time they experienced and this perception error influences their decision to travel.

In the psychological research field, Hornik's (1992) research focuses on the effect of affective moods as a situational variable on temporal judgement. It is hypothesized that mood biases temporal judgement by influencing the information to be recalled from memory. Individuals in a good mood, for instance, are prone to retrieve positive information, which in turn biases their judgement in a direction congruent with the mood. In transportation research, Bates, et al. (2001) 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 (2001) 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 (2000) 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 (Ben-Akiva, et al. 1999, 2002; Bolduc, et al. 2005; Walker & Ben-Akiva 2002). Integrated Choice and Latent Variable models are primarily employed for including attitudes and perceptions as explanatory variables of the choice, using psychometric scales as indicators of unobservable latent constructs (Atasoy, et al. 2011; Glerum, et al. 2011; Schüssler & Axhausen 2011). This methodology could potentially be used to deal with any type of variable which affects the choice. Walker, et al. (2010) 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 Integrated Choice and Latent Variable model for true travel time leads to significant shifts in both the travel time and travel cost variables, 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. (2013) 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 (Daly & Zachary 1977) or when applying multiple imputation (e.g. Brownstone & Steinmetz 2005) originally proposed by Rubin (1987). Multiple imputation can be used when accurate data for a subsample of the observations are available. Bhat (1994) 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 Integrated Choice and Latent Variable framework is proposed to correct measurement errors in a mode choice context (Walker, et al. 2010; Hess, et al. 2013).

The paper is structured as follows. Section 2 presents the transport mode choice case study, the statistical analysis of the dataset and the data processing procedure. Section 3 provides the methodology and includes the model specification regarding the Multinomial Logit model and the Integrated Choice and Latent Variable models. Section 4 presents the discussion of the estimation results obtained by using the extended software package BIOGEME (Bierlaire & Fetiarison 2009). Section 5 illustrates the validation and policy analysis. Section 6 provides the conclusions and suggestions for future research, while also discussing the limitations concerning the dataset used and possible extensions.

2. Survey and data collection

A comprehensive data collection campaign was carried out in Trieste between November 2009 and January 2010, within the framework of the UniMob project - the Mobility Management project for the University of Trieste. The purpose of this project was to study the travel behaviour of staff members and students in order to define the transport demand of the whole university population, both at local and regional levels.

The UniMob project was developed on the basis of a wide interdisciplinary approach thanks to the collaboration of four different departments at the University of Trieste, working on complementary research areas. The analysis of the university population together with the qualitative and the quantitative surveys was conducted by the Department of Civil Engineering and Architecture (DICAR).

First the university population was investigated in order to identify its socio-demographic characteristics, people's geographic origins and the distribution of users across faculties. The transport facilities available to university users were analysed, both at urban and extra urban levels. Following this analysis, the study area was selected: it included the province of Trieste which is covered by the local transport network and the regions Friuli – Venezia Giulia and Veneto, where railways and extra urban buses offer transport facilities which enable travellers to commute on a daily basis to Trieste.

The second step was to do a qualitative survey. It consisted of focus groups with students attending different faculties and involved interviewing teaching staff, technical administrative staff, organizers who define the lesson timetables and experts in the field of transportation. During this phase, several homogeneous groups that seem to differ in terms of their travel behaviours, availability of means and modal choices were identified according to the role occupied (teaching and technical administrative staff, students), the frequency of attendance (regular, occasionally) and the residence status (resident or domiciled in the city or the province of Trieste, residence outside the city and the province of Trieste).

The third step consisted of a revealed preference (RP) survey, which is the data source used in this research. The quantitative survey was performed through an on-line questionnaire: for this purpose, a software was programmed on the platform CAWI (Computer Assisted Web Interview). This choice allowed the whole university population (24.685 users) to participate in the interview. During November 2009 all the students regularly registered to the University (21.601 users) and all the teaching and administrative staff with a contract (3.084 users) were invited by email to complete the questionnaire. This procedure was preferred in order to minimise the distortion of the sample: all the users had the same opportunity to participate in the survey and this possibility was not affected by the frequency of being at the university. The quantitative questionnaire was designed following the outcomes of the preliminary phases mentioned above. Respondents reported 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 and arrival and departure times. Further data about travel habits, elements which impact modal choices, potential reasons for mode switching, perception of the risks associated each mode and opinions on topics related to urban mobility were gathered.

In total 3976 valid questionnaires were collected (response rate: 16,11%). In order to identify the factors driving individuals' mode choices over the reported sequences of trips, firstly all the information available is analysed and classified. Then, some preliminary descriptive statistics are elaborated in order to identify the socio-economic characteristics of the respondents, the transport modes available and the alternative chosen. The proportions of individuals in each category in the population and in the survey sample are reported in Table 1. It could be noted that in the sample some socio-demographic categories are oversampled, i.e. staff members and students residents in the province of Trieste.

Table 1 Proportions of socio-demographic categories (sample and population).

Category	Sample	Population
Role occupied		
Staff	26,94 %	12,49 %
Students	73,06%	87,51 %
Residence and role occupied		
Residents – Staff	77,96%	79,18%
Commuters - Staff	18,49%	16,96%
Others - Staff	3,55%	3,86%
Residents - Students	49,02%	31,02%
Commuters - Students	48,98%	66,41%
Others - Students	2,00%	2,57%

Since preliminary descriptive statistics show that differences between the outward voyage and the way back affect only 14% of the respondents, only the former is analysed. For the home - university trip, the main mode is identified by enumerating the sequence of means reported and the last mode chosen to reach the university.

The choice of the transportation mode is assumed to be among five alternatives: car, motorcycle, public transport and walk. All or part of these modes are available to each user, depending on their availability.

2.1 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. First the OD matrices are constructed, second the distances between each origin and destination reported are calculated, third the travel times for each alternative mode and the time necessary to find a parking lot are imputed, fourth the travel costs are calculated for the chosen and the unchosen alternatives.

Within the UniMob project, an assignment model was developed using the software Visum in order to analyse the public transport demand. In the assignment model, each zone was represented through a point, placed in the barycentre of the zone. Each trip was modelled as a travel between the barycentres of the corresponding origin-destination zones. The OD couples are assigned to each user, checking with great accuracy the correspondence between 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.

The distance between each origin and destination is calculated using the website Google Maps using the addresses of origin and destination reported by respondents.

The imputation of travel time is elaborated for each alternative mode separately, using different devices such as the assignment model made by Visum and Google Maps. When private modes (car and motorcycle) were used to access public transport (e.g. in sequences car-train-walk, motorcycle-train-bus, bike-extraurban bus), PT is considered as the chosen alternative and the travel time imputed is that of a trip travelled by public transportation between the corresponding origin and destination. The travel time by car is calculated using Google Maps, considering the origin and the destination reported by the respondents for each trip. In addition, in order to evaluate the real speed experienced by travellers during the morning peak hour, which is lower than the one calculated by the website, the values of travel time for car are assumed to be equal to the values calculated by Google Maps increased by 30%. The travel time by motorcycle is calculated using Google Maps, considering the origin and the destination reported by the respondent for each trip. Taking into account the specific cinematic characteristics of the motorcycles, faster than cars in the traffic flow, no penalizations are considered for the calculated travel time by the website. The travel time by PT is computed by the model of assignment created with the software Visum on this dataset. The period studied by the assignment model is between 6 and 10 am. Each trip is modelled as a travel between the barycentres of the corresponding origin-destination zones. For 672 users the travel time by PT is unknown, since these trips were not represented by the assignment model, due to a too limited number of travellers between the corresponding OD couples. The travel time by foot is calculated using the above-mentioned distances and assuming an average speed of 4 km/h, which corresponds to the average speed in urban areas given by the Highway Capacity Manual (2010).

Respondents who choose car reported the time they spent to find a parking lot when they arrived at the university. This information is processed to infer the average time necessary to find a parking lot for each user, depending on the faculty reached and the arrival time declared.

The imputation of costs for the chosen alternative is elaborated considering: an average cost of fuel calculated and the parking cost reported for respondents who choose car and motorcycle; the cost of the ticket reported for respondents who choose PT. The imputation of costs for the unchosen alternatives is elaborated considering: an average cost of fuel calculated and a parking cost equal to zero for car and motorcycle; the cost of a single trip ticket for PT.

The statistical analysis performed suggests that the sample is composed by groups which differ in terms of travel behaviour (i.e. staff/students, systematic/not systematic trips, urban/extra urban trips). Therefore, the sample heterogeneity is explored using heterogeneous-group Logit models which test the existence of scale parameters for different groups in the sample. Estimation results shows that the hypothesis that a scale parameter exists cannot be rejected between the following groups:

1. Staff and students;
2. Systematic and not systematic users;
3. Travellers performing urban and extra urban trips.

In this phase, the homogenous group of major concern that will be modelled in the forthcoming steps is chosen. Systematic trips completed inside the urban area represent the major part of the observations and are considered more relevant in order to analyse the travel behaviour of users. In addition, the observations reported by staff members appear to be more consistent than those reported by students. Due to the above-mentioned motivations, finally the sample selected includes 901 observations reported by staff members who systematically perform urban trips. For the sample selected, the alternatives available and the alternative chosen by respondents are reported in Table 2.

In the remainder of the present paper, only the sample selected in this section will be considered.

Table 2 Selected sample: alternative chosen and alternatives available.

Category	Car	Moto	PT	Walk
Alternative chosen	477	88	214	121
Alternative available	752	208	901	901

2.2 Travel time indicators

In order to investigate 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 an assignment software (Visum) for public transport and from Google Maps for car and motorcycle. For the chosen alternative three different indicators of travel time are available:

- Reported arrival and departure times (expressed in hours and minutes);
- Reported travel time intervals (multiple choice);
- Calculated travel time (imputed using the procedures described in section 2.1).

For the unchosen alternatives only the calculated travel time is available. The difference between the arrival and the departure time declared by the respondent is chosen as the most reliable indicator of the reported travel time. This measure is assumed to be the first indicator of the unobservable true travel time. In addition, the calculated travel time is assumed to be the second indicator of the unobservable true travel time. For the selected sample, some statistics related to the calculated travel time for the chosen and unchosen alternatives and the reported travel time for the chosen alternative are reported in Table 3.

Table 3 Selected sample: statistics on travel times.

Category	Car	Moto	PT	Walk
Calculated travel time				
Mean value [min]	12,95	9,58	19,16	71,49
Missing values [%]	-	-	14,21	-
Reported travel time				
Mean value [min]	23,09	14,67	31,17	21,13
Missing values [%]	-	-	-	-

In order to investigate the underlying nature of travel time perception, the travel time for each alternative mode is analysed separately, focusing on the gap existing between reported and calculated travel time. The reported travel time by respondents and the calculated travel time for the chosen mode are plotted separately for each alternative in a histogram in Figure 1 - Figure 4, rounding the values to the closest multiple of 1 minute. It could be noted that:

- The mean of the distribution of the reported travel time does not match the mean of the distribution of calculated travel time for each mode;
- The reported travel time is overestimated compared to the calculated travel time for all modes except walk;
- The density of reported travel time values is higher for multiples of 5, 15, 30 and 60 minutes.

Figure 1 Histogram of reported and calculated travel time for car [minutes].

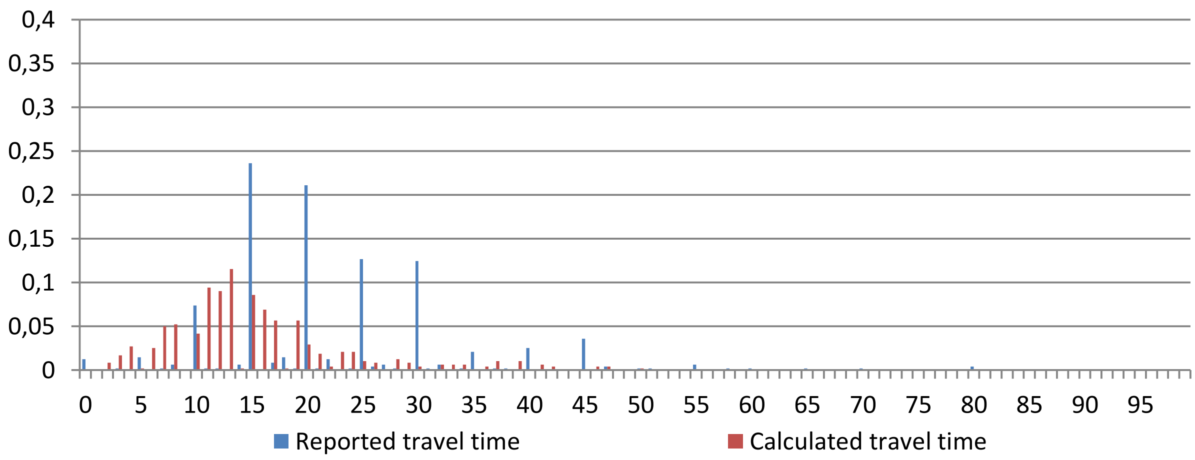


Figure 2 Histogram of reported and calculated travel time for motorcycle [minutes].

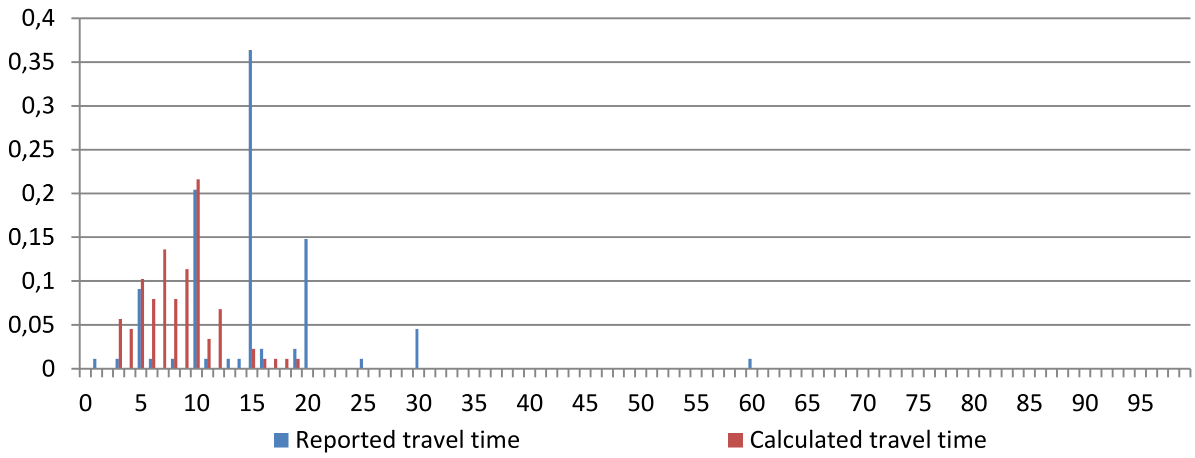


Figure 3 Histogram of reported and calculated travel time for PT [minutes].

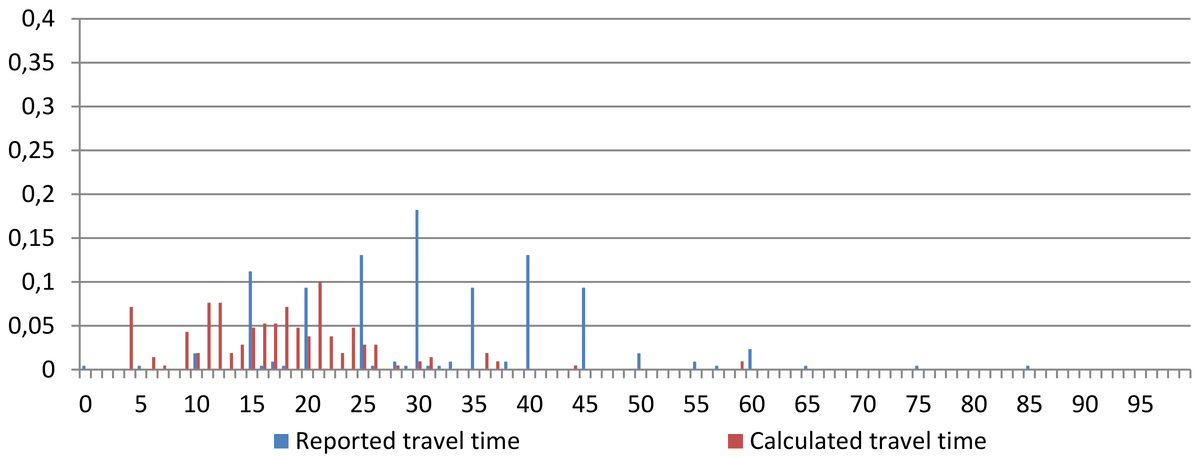
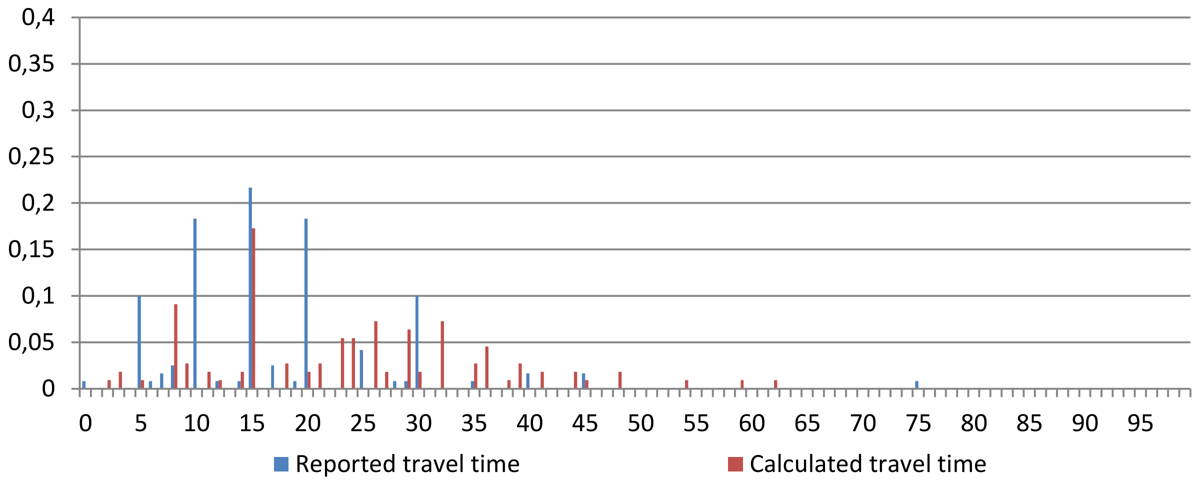


Figure 4 Histogram of reported and calculated travel time for walk [minutes].



3. Methodology

The aim of the present research is to investigate the impact of the underlying travel time perception of travellers on the modal choice. Therefore, the unobservable true travel time is modelled as a latent variable in order to understand how the error due to the inclusion of the calculated travel time into the mode choice model only influences the modal choice. For this purpose, the methodology proposed consists in the inclusion of different travel time indicators within the Integrated Choice and Latent Variable framework.

As a base reference a Multinomial Logit model is estimated, which has the same specification as the choice model included in the Integrated Choice and Latent Variable models. In the Multinomial Logit, the calculated travel time by different devices, such as Google Maps for car, motorbike and walk and by an assignment model for PT, is directly included into the utility function for each alternative mode. This model will be used as reference to evaluate the added value of the introduction of a latent attribute.

Then two Latent Variable models for true travel time are integrated into the discrete choice model, using the calculated travel time in the first one and the reported travel time in the second one as indicator of the true travel time. In the Integrated Choice and Latent Variable models the choice of the alternatives is assumed to be influenced by the effect of the latent attribute, alternative specific, which replaced the calculated travel time.

The calculated travel time is supposed to be affected by different errors related to the instrument employed for imputation. In order to model this phenomenon, a Latent Variable model should be added separately for each alternative mode. In the present research, the methodology proposed is employed to correct for the travel time of public transport because of two main reasons:

- The network level of service is expected to be lower for PT (i.e. aggregate travel time data, derived by an assignment model) than all the other modes (i.e. disaggregate travel time data, measured by Google Maps for each respondent);
- The gap between the reported travel time and the calculated travel time seem to affect the travellers who choose PT more than the travellers who choose the other alternatives.

3.1 Integrated Choice and Latent Variable framework: specification 1

In order to correct for a potential estimation bias in the estimation of the time parameter when the calculated travel time is directly used in the utility function, an Integrated Choice and Latent Variable model (ICLV) is estimated, assuming that the true travel time is a latent attribute and the calculated travel time could be used as an indicator.

The model schematized in Figure 5 is called the ICLV for the true travel time, since a latent attribute (the true travel time) is integrated into the choice model. The same framework was proposed by Walker, et al. (2010) in order to deal with measurement errors in the calculated travel time. Observed 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.

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*:

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

Where c and σ 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 σ. The distribution of the latent variable $TT*$ is f_1 $(TT_n^*; c, \sigma)$.

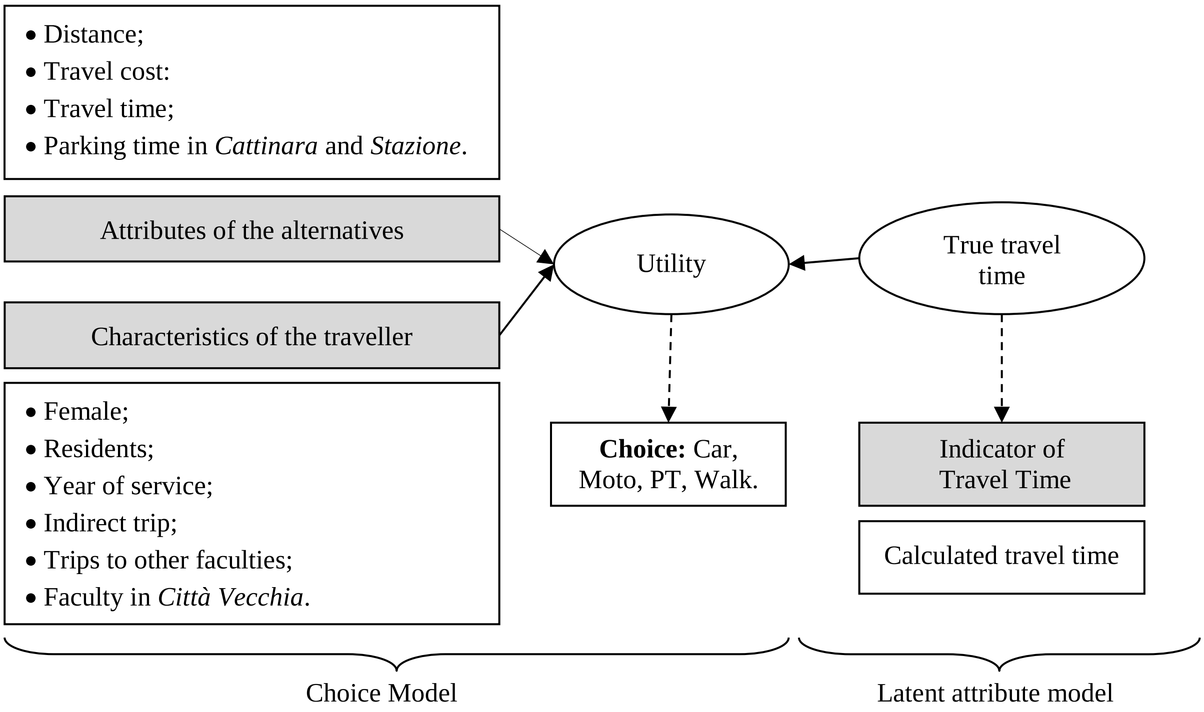
Latent Variable model: measurement equations for latent attribute

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:*

$$I_{1n} = \alpha_1 + \lambda_1 \cdot TT_n^* + \sigma_1 \cdot \delta_{1n}, \quad with \ \delta_{1n} \sim N(0, \sigma_1) \tag{2}$$

Where $\alpha_1$and λ_1 are parameters which are fixed for normalization purposes, σ_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 σ_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 $(I_{1n} | \ TT_n^*; \ \alpha, \lambda, \sigma)$.

Figure 5 Integrated Choice and Latent Variable framework for true travel time: specification 1.[1]



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_i, X_n and of the latent attribute TT_n^* as given in the equation (3):

$$U_{in} = V(X_i, X_n, TT_n^*; \beta) + \varepsilon_{in} \; with \; \varepsilon_{in} \sim EV(0,1) \quad (3)$$

Where

X_i is a vector representing the attributes of the alternative *i*;

[1] The Latent Variable model is similar to the one proposed by Walker, et al. (2010).

X_n is a vector representing the characteristics of the decision-maker;

TT_n^* is the unobservable true travel time;

β is a vector of parameter to estimate;

ε_{in} is the error term.

Integrated model framework

In the contest of discrete choice, the probability that the individual *n* chooses the alternative *i* is given by equation (4):

$$P(y_{in} = 1) = \Pr\left(U_{in} \geq U_{jn}, \forall\, j\right) \tag{4}$$

Under the assumption that the error terms are independent, the likelihood function is given by the formula (5):

$$\mathcal{L}_n\left(y_{n,} I_1 \middle| X_i, X_n;\ \alpha, \lambda, \beta, \sigma\right) =$$

$$\int_{TT^*} P(y_n\,|X_i, X_n, TT_n^*;\ \beta) \cdot\ f_2\ (I_{1n}|\ TT_n^*;\ \alpha, \lambda, \sigma)\ \cdot f_1\ (TT_n^*;\ c, \sigma)\, dTT^* \tag{5}$$

Where

$f_2\ (I_{1n}|\ TT_n^*;\ \alpha, \lambda, \sigma)$ 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 (6):

$$\max_{\alpha,\lambda,\beta,\sigma} \sum\nolimits_{\mathrm{n}} \log\left(\mathcal{L}_n\left(y_{n,} I_1 \middle| X_i, X_n;\ \alpha, \lambda, \beta, \sigma\right)\right) \tag{6}$$

The specification is reported in Table 4 and the explanatory variables used in the utilities of the choice model are listed as follows:

- *Staff* is a dummy variable equal to 1 if status = staff;
- TT_{CAR_STAFF} / D_{STAFF} represents the travel time by car in minutes divided by the total distance travelled;
- TT_{MOTO_STAFF}, TT_{PT_STAFF}, TT_{WALK_STAFF} represent the travel time in minutes for each mode;
- $MissingTime_{\mathrm{STAFF}}$, is a variable which assumes values equal to 1 when the PT travel time is missing;
- $C_{CAR\_FUEL\_STAFF}$, $C_{MOTO\_FUEL\_STAFF}$, C_{PT_STAFF} are the travel costs in euros;
- $ParkingTime_{STAFF}$ represents the parking time function built as described in chapter 2.1, referring to the faculties located near Ospedale Maggiore and near the railway station; the parking time has no impact for trips directed to any of the other faculties;
- $Female_{STAFF}$ is a dummy variable being 1 for respondents who are women and staff;
- $YearServ_{STAFF} \cdot (YearServ_{STAFF}>20)$ is a piecewise linear variable equal to 0 if the years of service is lower than 20 and equal to the year of service otherwise; therefore individuals who reported a number of year of service lower than 20 constitute a reference value and the parameter is estimated for the remaining population;
- $IndirectTrip_{STAFF}$ is a dummy variable equal to 1 for respondents who reported to have stopped once or more for different purposes during their home – university trip;
- $OtherFaculties_{STAFF}$ is a dummy variable equal to 1 for respondents who reported to travel during the day between two or more faculties;
- $CittàVecchia_{STAFF}$ is a dummy variable equal to1 for respondents who perform a trip to one of the faculties located in Città Vecchia.

Several cost distributions are tested and those which fit the observations best are selected:

- When PT is the chosen alternative, cost is assumed to be equal to the ticket cost for respondents who hold a ticket and equal to zero for respondents who hold a pass;
- When PT is not the chosen alternative, cost is assumed to be equal to the cost of a ticket for everyone;
- For car and motorcycle, the cost is assumed to be equal to the cost of fuel for each user.

Table 4 Integrated Choice and Latent Variable Framework for true travel time: specification 1. Specification table of the utilities and latent attribute model.

Utilities	V_{CAR}	V_{MOTO}	V_{PT}	V_{WALK}
ASC_{CAR_STAFF}	Staff	-	-	-
ASC_{MOTO_STAFF}	-	Staff	-	-
ASC_{PT_STAFF}	-	-	Staff	-
β_{COST_STAFF}	$C_{CAR\_FUEL\_STAFF}$	$C_{MOTO\_FUEL\_STAFF}$	C_{PT_STAFF}	-
$\beta_{TIME\_CAR\_STAFF}$	TT_{CAR_STAFF} /D_{STAFF}	-	-	-
$\beta_{TIME\_MOTO\_STAFF}$	-	TT_{MOTO_STAFF}	-	-
$\beta_{TIME\_PT\_STAFF}$	-	-	TT*	-
$\beta_{TIME\_WALK\_STAFF}$	-	-	-	TT_{WALK_STAFF}
$\beta_{MISSING\_TIME\_PT\_STAFF}$	-	-	-	-
$\beta_{PARKING\_CAR\_STAFF}$	$ParkingTime_{STAFF}$	-	-	-
$\beta_{GENDER\_MOTO\_STAFF}$	-	$Female_{STAFF}$	-	-
$\beta_{YEAR\_CAR\_STAFF}$	$YearServ_{STAFF} \cdot (YearServ_{STAFF}>20)$	-	-	-
$\beta_{INDIRECT\_TRIP\_CAR\_STAFF}$	$IndirectTrip_{STAFF}$	-	-	-
$\beta_{FACULTIES\_PT\_STAFF}$	-	-	$OtherFaculties_{STAFF}$	-
$\beta_{CITTAVECCHIA\_WALK\_STAFF}$	-	-	-	$CittaVecchia_{STAFF}$

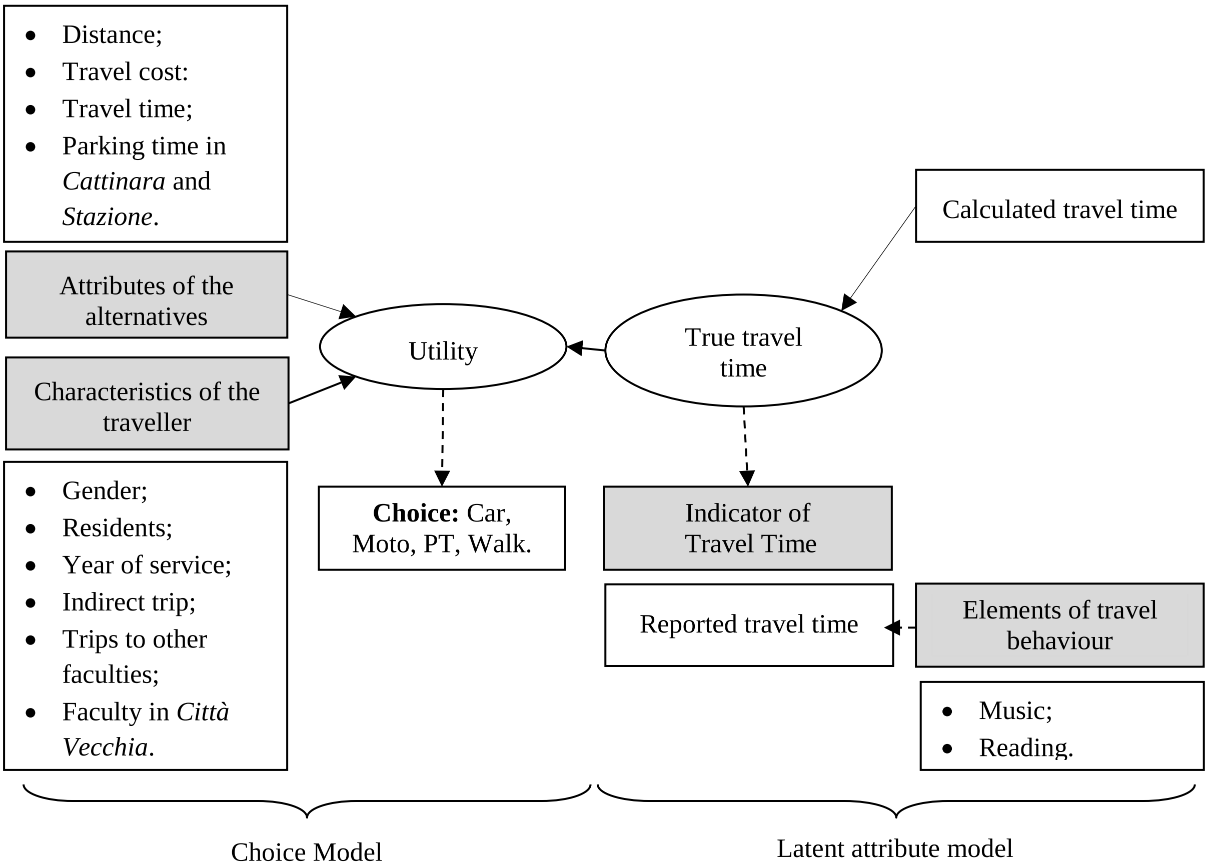
Latent attribute model				
Latent time – c	-	-		-
Latent time – λ	-	-		-
Latent time – σ	-	-		-
Meas. Equ. – σ_1	-	-		-

3.2 Integrated Choice and Latent Variable framework: specification 2

In order to introduce the reported travel time by respondents into the choice model instead of the calculated travel time thereby correcting for a potential estimation bias in the estimation of the time parameter, an Integrated Choice and Latent Variable model is estimated, assuming that the true travel time is a latent attribute. It is important to note that the model schematized in Figure 6 differs from the specification 1 proposed as follow:

- The reported travel time is used as an indicator of the true travel time in the measurement equation;
- The calculated travel time is included in the structural equation, assuming that it affects the true travel time;
- Elements of travel behaviour are included in the measurement equation, assuming that they influence the reported travel time.

Figure 6 Integrated Choice and Latent Variable framework 2 for true travel time: specification 2.



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 (7) for the public transport alternative and each individual *n*:

$$TT_n^* = c + \lambda \cdot TT_n + \sigma \cdot \delta_n, \; with \; \delta_n \sim N(0,1) \tag{7}$$

Where TT_n is the calculated travel time, c, λ and σ are parameter to be estimated.

The distribution of the latent variable TT_n^* is $f_1 \left(TT_n^* | TT_n; \; c, \; \lambda, \; \sigma\right)$.

Latent Variable model: measurement equations 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 (8):

$$I_{2n} = \alpha_2 + \lambda_2 \cdot TT_n^* + \beta_2 \cdot X_n + \sigma_2 \cdot \delta_{2n}, \quad with \; \delta_{2n} \sim N(0, \sigma_2) \tag{8}$$

where α_2 and λ_2 are parameters which are fixed for normalization purpose, β_2 and σ_2 are parameters to be estimated, TT_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 $f_2 \left(I_{2n} | \, X_n, TT_n^*; \; \alpha, \lambda, \beta, \sigma\right)$.

Some parameters of the model are normalized for identification purposes. Firstly α_2 is fixed to 0 and λ_2 is fixed to 1, secondly $c, \lambda, \sigma, \sigma_2$ are estimated.

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 is performed with the whole set of elements of travel behaviour as an exploratory step. Principal Component Analysis (PCA) is a method which analyses how unobservable constructs could influence a measured variable (i.e. reported travel time) by examining the set of correlations between the observed variables (i.e. elements of travel behaviour). The principal component analysis is performed employing the statistical software R, using the package *psych* version 1.2.8 developed by Revelle (2012).

The respondents who reported in the questionnaire that they usually choose public transport are selected and the principal component analysis is performed on this group. In Table 5 results are presented for the first three components having a loading higher than 0,3 (in absolute sense). The three components identified could be defined as follows: *short trips*, *long trips*, *middle trips*. Analysing the results it could be noted that the first component is negatively correlated with the value of calculated travel time (i.e. short trips), positively correlated with the overestimation of reported travel time compared to calculated travel time (i.e. overestimation is frequent) and to the habit to not listen to music during the home–university trip (i.e. listening to music rarely). The second component is positively correlated with the value of calculated travel time (i.e. long trips) and with the gap between reported travel time and calculated travel time, negatively with the habit of reading for leisure during the home–university trip (i.e. reading for leisure is frequent). The third component is positively correlated with the value of reported travel time, with the gap between reported travel time and calculated travel time, with the importance of comfort in the modal choice and with the perception that the quality of public transport service being higher in Trieste than in the respondent's home town.

Table 5 Principal component analysis[2] for elements of travel behaviour – Habitual choice PT.

Elements of travel behaviour and travel time perception	*Short trips*	*Long trips*	*Middle trips*
I never listen to music during the trip home-university.	0,389		
I never read for leisure during the trip home-university.		-0,671	
The comfort of the whole trip affects the modal choice.			0,387
PT in Trieste is better than PT in my home town.			0,897
Calculated travel time.	-0,553	0,672	
Reported travel time with arrival/departure time.		0,774	0,385
Gap between the reported travel time with arrival/departure time and the calculated travel time.	0,692	0,339	0,390
Overestimation of the reported travel time with arrival/ departure time compared to the calculated travel time.	0,896		

[2] Parallel Analysis (Horn 1965) is performed in order to identify how many dimensions to use to represent the correlation matrix. In addition, pair-wise deletion is introduced to deal with missing values in the correlation matrix and the rotation "varimax" is selected.

The elements of travel behaviour introduced in the measurement equation are selected referring to the PCA presented above. The introduction of each variable is tested using likelihood ratio tests in order to define if the improvement is significant or not. The element related to the superior PT quality in Trieste comes out not to be significant because of the high number of missing answers.

The explanatory variables used in the measurement equation are listed as follows:

- $Reading_{STAFF}$ is a variable corresponding to the statement "I never read during the trip home-university" (five-point Likert scale);
- $Music_{STAFF}$ is a variable corresponding to the statement "I never listen to music during the trip home-university" (five-point Likert scale);
- $MissingRead_{STAFF}$ is a variable which assumes values equal to 1 when the variable $Reading_{STAFF}$ is missing ;
- $MissingMusic_{STAFF}$ is a variable which assumes values equal to 1 when the variable $Music_{STAFF}$ is missing.

Discrete choice model and integrated model framework

The discrete choice model and the integrated model framework are analogous to those proposed in section 3.1. The specification is reported in Table 6 and the explanatory variables used in the utilities are listed in section 3.1.

Table 6 Integrated Choice and Latent Variable Framework for *true travel time*: specification 2. Specification table of the utilities.

Utilities	V_{CAR}	V_{MOTO}	V_{PT}	V_{WALK}
ASC_{CAR_STAFF}	Staff	-	-	-
ASC_{MOTO_STAFF}	-	Staff	-	-
ASC_{PT_STAFF}	-	-	Staff	-
β_{COST_STAFF}	$C_{CAR\_FUEL\_STAFF}$	$C_{MOTO\_FUEL\_STAFF}$	C_{PT_STAFF}	-
$\beta_{TIME\_CAR\_STAFF}$	TT_{CAR_STAFF} /D_{STAFF}	-	-	-
$\beta_{TIME\_MOTO\_STAFF}$	-	TT_{MOTO_STAFF}	-	-
$\beta_{TIME\_PT\_STAFF}$	-	-	TT*	-
$\beta_{TIME\_WALK\_STAFF}$	-	-	-	TT_{WALK_STAFF}
$\beta_{MISSING\_TIME\_PT\_STAFF}$	-	-	-	-
$\beta_{PARKING\_CAR\_STAFF}$	$ParkingTime_{STAFF}$	-	-	-
$B_{FEMALE\_MOTO\_STAFF}$	-	$Female_{STAFF}$	-	-
$\beta_{YEAR\_CAR\_STAFF}$	$YearServ_{STAFF} \cdot (YearServ_{STAFF}>20)$	-	-	-
$\beta_{INDIRECT\_TRIP\_CAR\_STAFF}$	$IndirectTrip_{STAFF}$	-	-	-
$\beta_{FACULTIES\_PT\_STAFF}$	-	-	$OtherFaculties_{STAFF}$	-
$\beta_{CITTAVECCHIA\_WALK\_STAFF}$	-	-	-	$CittaVecchia_{STAFF}$

Latent attribute model

Latent time – c	-	-		-
Latent time – λ	-	-		-
Latent time – σ	-	-		-
$\beta_{READING\_STAFF}$	-	-	$Reading_{STAFF}$	-
β_{MUSIC_STAFF}	-	-	$Music_{STAFF}$	-
$\beta_{MISSING\_READING\_MUSIC}$	-	-	$MissingRead_{STAFF}$	-
Meas. Equ. – σ_1	-	-		-

4. Estimation results

The maximum likelihood method is used for model estimation, which is done by using the extended software package BIOGEME (Bierlaire & Fetiarison 2009). The specifications presented in chapter 3 are the best reached. There are two ways to estimate the integrated model: the sequential and the simultaneous approach. The sequential estimation method involves first estimating the Latent Variable model using standard latent variable estimators. The second step is to use fitted latent variables and their distributions to estimate the choice model, in which the choice probability is integrated over the distribution of the latent variables.

In this research the two Latent Variable models proposed in chapter 3 for the PT alternative are estimated separately. Several specifications are tested in order to choose the variables that affect the reported travel time. Then a residual analysis is performed in order to find the best one. This step is followed by the creation of the ICLVs, estimated sequentially. The estimation results are presented in Table 8. The first column reports the results of the base model (i.e. Multinomial Logit model estimated for staff members), the second reports the results of the ICLV for true travel time specification 1 and the third reports the results of the ICLV for true travel time specification 2.

It is important to note that the Multinomial Logit model estimated reproduce correctly the choice probabilities of that number of observations, while the Integrated Choice and Latent Variable models do not. In order to reproduce the real market shares with the ICLV, the alternative specific constants have to be adjusted. The integrated models are simulated a first time and then the constants are corrected for each mode using the formula (9) (Ben-Akiva & Lerman 1985):

$$a_{k+1,i} = a_{k,i} + \ln\left(\frac{P_{0,i}}{P_{k,i}}\right) \tag{9}$$

Where:

$a_{k+1,i}$ is the corrected alternative specific constant for the mode i;

$a_{k,i}$ is the alternative specific constant calculated in the previous step for the mode i;

$P_{0,i}$ is the market share for the mode i calculated considering the choices reported by respondents;

$P_{k,i}$ is the market share for the mode i reproduced by the model in the previous step.

After the correction of the constants, the Integrated Choice and Latent Variable models are simulated a second time. The procedure is iterated until the difference between the market shares observed and the values predicted by the model is considered acceptable. In the present research, it is assumed that an acceptable difference is equal to 0,05%.

The log-likelihood and the goodness of fit results are reported in Table 7. The Final log-likelihood values and the values of rho-bar-squared are calculated for only the choice component of the ICLVs to be comparable with the base Multinomial Logit model. The number of parameters estimated J and the number of parameters estimated for the choice component K are reported for each model.

It could be noted that the fit of the choice component of the ICLVs for true travel time and the likelihood function decrease over the Logit model. These results are consistent with recent findings by Vij & Walker (2012). Indeed, they discovered 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.

In addition, the ICLV specification 2 has the best fit compared to the ICLV specification 1. This result indicates that the second specification proposed seems to provide a richer behavioural explanation of mode choices.

Looking at the utility parameters of the Multinomial Logit 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.

Table 7 Statistics for the Integrated Choice and Latent Variable models.

Statistics	Base model	ICLV specification 1	ICLV specification 2
J	15	18	21
K	15	15	14
Number of observations	901	901	901
Final log-likelihood	-422,448	-490,554	-438,501
Adjusted rho-bar-squared	0,559	0,491	0,544

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 the car alternative, it is important to point out that 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_{CAR}/D is introduced into the utility function. 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. In fact, the coefficient $\beta_{YEAR\_CAR\_STAFF}$ has a positive sign and a staff member whit more than 20 years of service has a higher probability of choosing car. 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;
- 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 one or more 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;
- Staff members who work in a faculty located in Città Vecchia (i.e. the old city centre) prefer walking to all other modes. Possible explanations could be that users have a preference for walk in the city centre and that in the city centre residences are located closer to the faculties; in addition it could be noted that those faculties are located in

the lower part of the city, where the ground is almost flat, resulting in a lower effort to walk.

The estimated parameters of the Logit model and the choice component of the ICLV models are close to each other, with the exception of the time parameters. Walker, et al. (2010) note significant shifts in the time and cost parameters of the ICLV model estimated and point out that these results indicate a correlation between the time and cost variables. Looking at the time parameter of PT, that is associated to the latent time estimated, the magnitude increases in both ICLVs compared to the Logit model.

Looking at the Latent Variable models, the means of the latent travel times are expected to have a magnitude and sign which can be compared to the values of the reported travel time and of the calculated travel time present in the raw dataset.

Referring to the ICLV for true travel time specification 1, the mean of the latent time c is equal to 19,50 minutes and the standard deviation σ 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 σ_1 equal to 2,420 minutes. This standard deviation is significant, indicating that there is a measurement error inherent in the network derived travel times.

Referring to the ICLV for true travel time specification 2, the mean of the latent time c is equal to 22,40 minutes, the calculated travel time affects the latent time significantly with a magnitude λ equal to 0,587 and the standard deviation σ is 1,320 minutes. The reported travel time is assumed to be normally distributed with mean equal to the latent travel time and standard deviation σ_1 equal to 2,390 minutes. Moreover the parameters $\beta_{READING\_STAFF}$ and β_{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 Principal Component Analysis. Clear conclusions for missing values cannot be drawn because the coefficient $\beta_{MISSING\_READING\_MUSIC}$ has a low t-statistic. Nevertheless missing values are kept in the measurement equation separately from non-missing values in order to distinguish their effect on travel time perception. 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 seem to be an overestimation of the calculated travel time, in accordance with the results obtained through statistical analysis.

Table 8 Integrated Choice and Latent Variable models: estimation results.

Parameters	Base model		ICLV specification 1		ICLV specification 2	
	Estimate	T-test	Estimate	T-test	Estimate	T-test
ASC_{CAR_STAFF}	1,130	1,70 *	2,284	-	1,290	2,08
ASC_{MOTO_STAFF}	0,295	0,60 *	0,150	-	0,079	0,16
ASC_{PT_STAFF}	3,120	5,64	7,730	-	6,860	4,38
ASC_{SM_STAFF}	-	-	-0,170	-	-	-
β_{COST_STAFF}	-4,690	-12,08	-4,690	-8,65	-4,750	-11,43
$\beta_{TIME\_CAR\_STAFF}$	-0,840	-4,78	-1,160	-6,85	-0,927	-5,83
$\beta_{TIME\_MOTO\_STAFF}$	-0,155	-3,66	-0,110	-2,43	-0,129	-3,05
$\beta_{TIME\_PT\_STAFF}$	-0,114	-7,50	-0,347	-4,08	-0,178	-4,59
$\beta_{TIME\_WALK\_STAFF}$	-0,085	-9,33	-0,082	-7,72	-0,085	-8,96
$\beta_{MISSING\_TIME\_PT\_STAFF}$	-5,420	-6,85	-	-	-	-
$\beta_{PARKING\_CAR\_STAFF}$	-0,292	-4,69	-0,291	-4,71	-0,252	-4,40
$B_{FEMALE\_MOTO\_STAFF}$	-0,960	-2,65	-0,974	-2,64	-0,936	-2,62
$\beta_{YEAR\_CAR\_STAFF}$	0,020	2,58	0,018	2,22	0,018	2,35
$\beta_{INDIRECT\_TRIP\_CAR\_STAFF}$	1,470	5,18	1,630	5,31	1,590	5,48
$\beta_{FACULTIES\_PT\_STAFF}$	0,815	3,18	0,616	2,12	0,509	2,01
$\beta_{CITTAVECCHIA\_WALK\_STAFF}$	0,853	2,14	0,664	1,56 *	0,848	2,19
Latent time – c	-	-	19,466	42,36	22,445	5,69
Latent time – λ	-	-	-	-	0,587	6,26
Latent time – σ	-	-	-0,551	-3,79	-1,322	-1,08 *
$\beta_{READING\_STAFF}$	-	-	-	-	-7,900	-2,69
β_{MUSIC_STAFF}	-	-	-	-	6,245	2,16
$\beta_{MISSING\_READING\_MUSIC}$	-	-	-	-	-2,199	-0,20 *
Meas. Equ. – σ_1	-	-	2,425	76,53	2,390	34,68

(* statistical significance <95%)

4.1 Residual analysis

Residual analysis is commonly performed to investigate regression models and it can also be used to assess the goodness-of-fit of the measurement equations of a Latent Variable model. For each estimated LVM, the residuals of the measurement equation will be analysed as a diagnostic of the normality assumption of the error term.

LVM for true travel time: specification 1

For the indicator I_{1n} of calculated travel time, the residuals are calculated as presented in the equation (10):

$$\delta_{1n} = I_{1n} - \alpha_1 - \lambda_1 \cdot TT^*_{n\_SIM} \tag{10}$$

Where $TT^*_{n\_SIM}$ are the fitted values of TT^*_n.

For the equation presented above, the corresponding Q-Q plot is reported in Figure 7. The Q-Q plot shows discontinuity both in the upper and the lower tail, and continuity in the centre. The discontinuity is due to the methodology used to impute calculated travel time for PT: the same value of travel time is associated to all users who perform a trip between the same zones of origin and destination, meaning that actually the calculated travel time for PT is not a continuous variable.

LVM for true travel time: specification 2

For the indicator I_{2n} of reported travel time, the residuals are calculated as presented in the equation (11):

$$\delta_{2n} = I_{2n} - \alpha_2 - \lambda_2 \cdot TT^*_{n\_SIM} - \beta_2 \cdot X_n \tag{11}$$

Where $TT^*_{n\_SIM}$ are the fitted values of TT^*_n.

For the equation above, the corresponding Q-Q plot is reported in Figure 9. In addition, a new model is estimated, that is similar to the previous LVM but does not contain the elements of travel behaviour in the measurement equation. In order to check the effect of the introduction of the above-mentioned elements in the measurement equation, the Q-Q plot of the residuals referring to the measurement equation without additional variables X_n is reported in Figure 8. Both Q-Q plots show a straight continuous line, suggesting first that the measurement equation proposed fits the data well and second that the introduction of elements of travel behaviour X_n in the measurement equation affects the residuals distribution just slightly.

Figure 7 Normal Q-Q plot of standardized residuals - LVM for true travel time: specification 1.

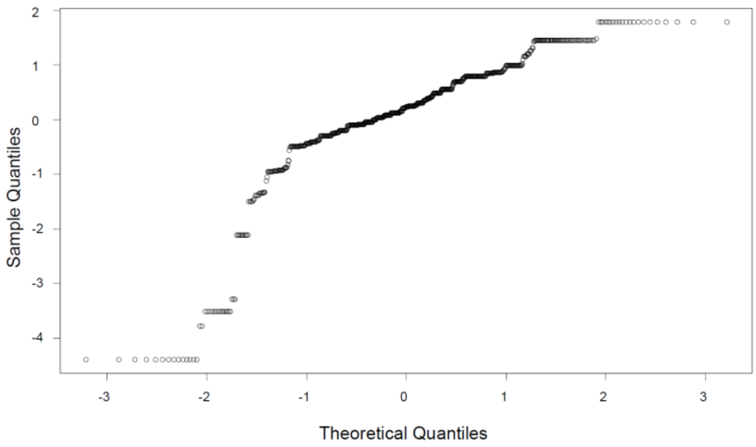


Figure 8 Normal Q-Q plot of standardized residuals - LVM for true travel time: specification 2 without elements of travel behaviour X_n.

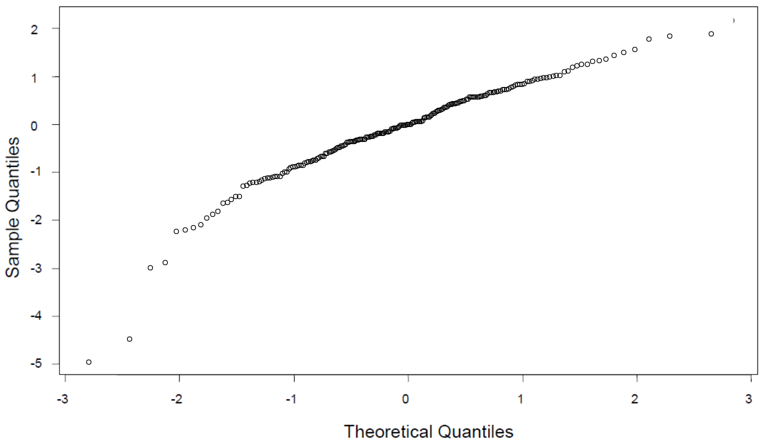
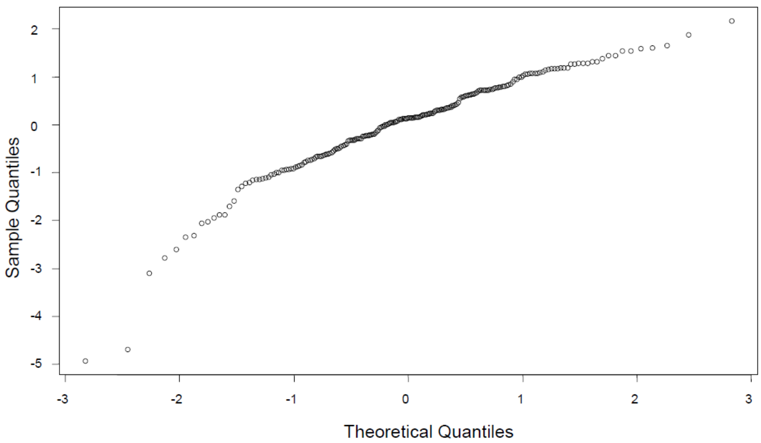


Figure 9 Normal Q-Q plot of standardized residuals - LVM for true travel time: specification 2.



5. Validation and policy analysis

The estimation results presented in chapter 4 can be used to quantify the demand by defining several indicators. 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 expresses the willingness to pay of individuals to gain a travel duration of one hour is computed for each alternative mode.

5.1 Validation analysis

A proper validation of the model would require its application on a different data set but no other similar dataset is available. As consequence, the dataset available is split into two parts.

First, 70% of the observations are selected randomly and the model is estimated on the latter. It is important to note that the Multinomial Logit model estimated on part of the observations reproduce correctly the choice probabilities of that number of observations, while the Integrated Choice and Latent Variable models do not. In order to reproduce the real market shares with the integrated models, the alternative specific constants are adjusted a second time for each mode using the formula proposed in chapter 4 (Ben-Akiva & Lerman 1985).

Second, the models are applied on the remaining 30% of the observations. Histograms of the choice probabilities predicting the transport mode choice chosen by the individuals in the 30% of the observations are shown for the Multinomial Logit model and for the Integrated Choice and Latent Variable models in Figure 10 – Figure 12. In addition, the average number of alternatives available for each respondent and the corresponding chance level are calculated. Table 9 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. It is important to note that for same observations the choice probabilities predicted by the ICLV specification 1 are higher than those predicted by the Multinomial Logit model.

Table 9 Percentages of choice probabilities higher than 0,33, 0,50, 0,70 and 0,90.

Threshold	Base model	ICLV specification 1	ICLV specification 2
33 %	84,21 %	80,07 %	83,08 %
50 %	74,43 %	72,55 %	72,93 %
70 %	60,53 %	61,27 %	59,77 %
90 %	37,59 %	43,98 %	37,59 %

Figure 10 Base model: percentages of the choice probabilities.

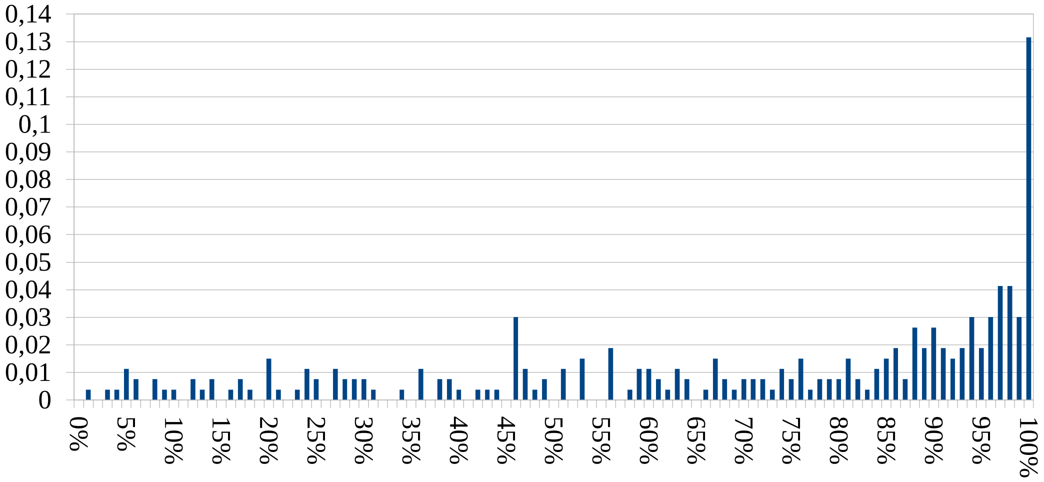


Figure 11 ICLV specification 1: percentages of the choice probabilities.

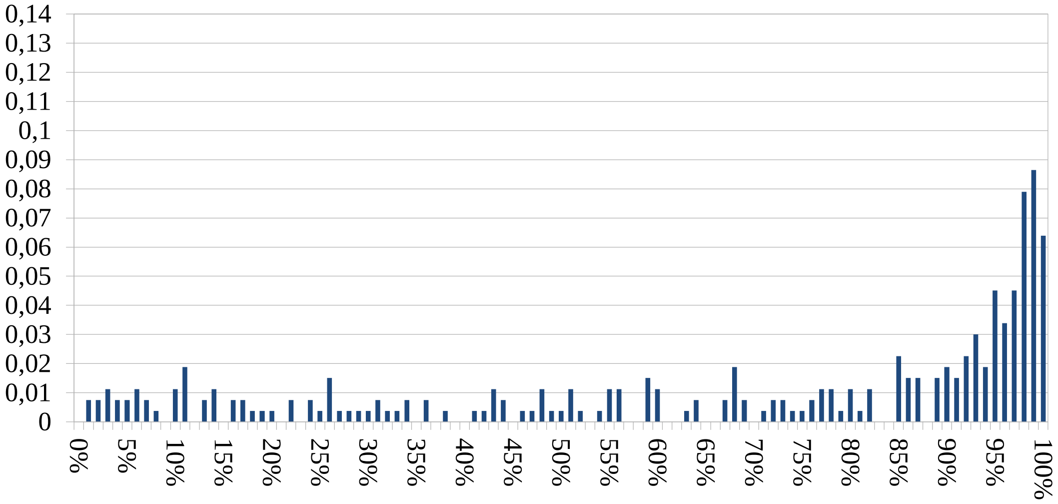
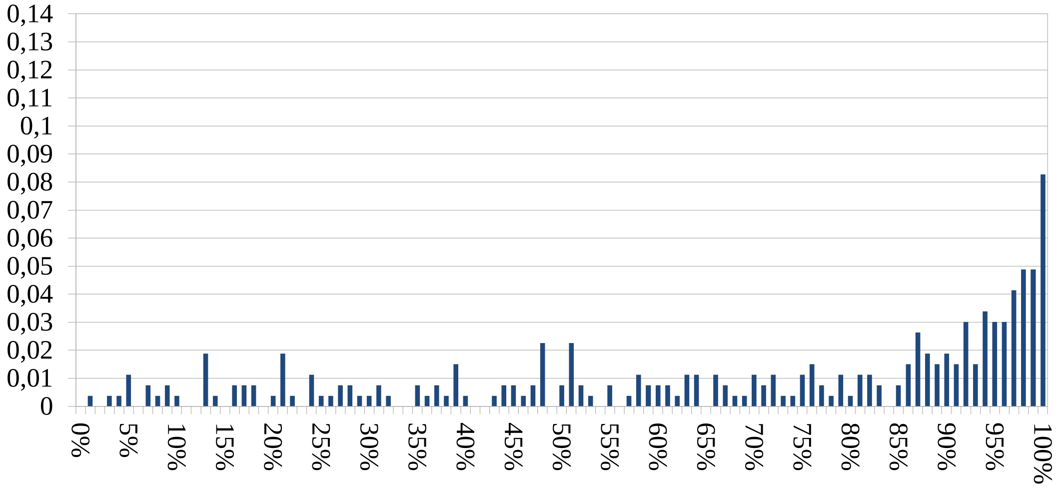


Figure 12 ICLV specification 2: percentages of the choice probabilities.



5.2 Value of time

The value of time (VOT) is an indicator of the willingness to pay (WTP) of individuals to reduce the duration of their trip by one hour. Since minutes is the unit of travel time, the disaggregate VOT for an individual *n*, for a decrease of 1 hour is computed as given in the equations (12) - (15):

$$VOT_{CAR\_STAFF,n} = \frac{\beta_{\mathrm{TIME\_CAR\_STAFF}} \cdot 60}{\beta_{\mathrm{COST\_STAFF}} \cdot \mathrm{D}_{\mathrm{STAFF}}} \quad (12) \qquad VOT_{MOTO\_STAFF,n} = \frac{\beta_{\mathrm{TIME\_MOTO\_STAFF}} \cdot 60}{\beta_{\mathrm{COST\_STAFF}}} \quad (13)$$

$$VOT_{PT\_STAFF\ ,n} = \frac{\beta_{\mathrm{TIME\_PT\_STAFF}} \cdot 60}{\beta_{\mathrm{COST\_STAFF}}} \quad (14) \qquad VOT_{PARK\_STAFF,n} = \frac{\beta_{\mathrm{PARKING\_CAR\_STAFF}} \cdot 60}{\beta_{\mathrm{COST\_STAFF}}} \quad (15)$$

Where:

β_{STAFF} are the parameters estimated;

$\mathrm{D}_{\mathrm{STAFF}}$ are the distances travelled.

The aggregate indicators of VOT are equal to the mean of the disaggregate VOT calculated and are reported in Table 10.

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.

Table 10 Value of time.

	Car	Moto	PT	Parking search time
Base model	4,58 €/h	1,98 €/h	1,46 €/h	3,74 €/h
ICLV specification 1	6,27 €/h	1,40 €/h	4,43 €/h	3,71 €/h
ICLV specification 2	4,94 €/h	1,62 €/h	2,25 €/h	3,19 €/h

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 (de Jong & Gunn 2001; Fiorello & Pasti 2003; Cherchi 2003; Catalano, et al. 2008). In addition, Rotaris, et al. (2012) estimate the VOT for students enrolled at the University of Trieste, developing a methodology which combines revealed and stated preferences: the VOTs vary from 1,4 to 2,8 €/h. The VOT obtained for PT seems to be lower than expected.

Referring to the Integrated Choice and Latent Variable 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 public transport shifts and the estimated VOT increases significantly, exactly as in the case study presented by Walker, et al. (2010). In the ICLV for true travel time specification 1, the value of time increases by over 200%, from 1,46 €/h to 4,44 €/h. In the ICLV for true travel time specification 2, the value of time increases by over 55%, from 1,46 €/h to 2,25 €/h.

In addition, the VOTs calculated for the other modes with the ICLVs change as follows:

- It increases for car, from 4,58 €/h to 6,27 €/h (ICLV specification 1) and 4,94 €/h (ICLV specification 2);
- It decreases for motorcycle, from 1,98 €/h to 1,40 €/h (ICLV specification 1) and 1,62 €/h (ICLV specification 2);
- It decreases for parking search time, from 3,74 €/h to 3,71 €/h (ICLV specification 1) and 3,19 €/h (ICLV specification 2).

The VOTs obtained for PT within the Hybrid Choice framework seem to be more consistent with the referential values for the VOT in Italy found in literature compared to the values obtained for PT with the Multinomial Logit model. Cherchi (2003) points out that using the parameters estimated with a Multinomial Logit, the VOT are largely underestimated and more realistic results could be obtained accounting for the variations in sensitivity among respondents. Fiorello & Pasti (2003) 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 for true travel time specification 1. Cherchi (2003) 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 for true travel time specification 2.

6. Conclusion and future research

The aim of the present research is to develop methods for discrete choice modelling that use available data and account for data limitations. In order to deal with measurement errors in travel time, the use of the hybrid choice framework proposed by Walker (2001) and Walker, et al. (2010) 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.

First, the transport mode choice case study is thoroughly analysed using statistics and the homogenous group that will be modelled is selected (i.e. staff, systematic trips, urban trips). Second, the data processing procedure presents the imputation of calculated travel time for each mode, using different devices such as an assignment model developed by Visum for public transport and Google Maps for car and motorcycle. Third, using the extended software package BIOGEME (Bierlaire & Fetiarison 2009), a Multinomial Logit is implemented. Fourth, two Latent Variable models for the value of travel time assumed to affect mode choices (i.e. "true travel time") are integrated into the discrete choice model. The methodology is employed to correct for the travel time of public transport, because of the 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" seem to affect more the travellers who choose PT than all the other alternatives. In the first Latent Variable model, the calculated travel time is used as an indicator of the true travel time. In the second Latent Variable model, the reported travel time is used as an indicator of the true travel time, assuming that the reported time is affected by elements of travel behaviour such as the habits of listening to music and reading during the home-university trip. The alternative specific constants of the Integrated Choice and Latent Variable models (ICLVs) have to be adjusted to reproduce the market shares. The estimation results obtained by the Multinomial Logit 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.

In addition, the Logit model and the ICLVs are validated in order to assess potential improvements in the forecasting power. The choice probabilities are well predicted by the Multinomial Logit model as well as by the ICLVs. It is important to note that for same observations the choice probabilities predicted by the ICLV specification 1 are higher than those predicted by the Multinomial Logit model.

Nevertheless further analysis is necessary, the methodology implemented is applicable to compute value of time for individuals. Results indicate that the Logit model which does not correct for measurement errors seem to underestimate travellers' value of time (1,46 €/h). In addition, the value of time computed using the first specification (4,44 €/h) is higher than the value of time computed using the second specification (2,25 €/h). In comparison to the Logit, 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 (Cherchi 2003) and 3,00-4,00 €/h (Fiorello & Pasti 2003).

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 ICLV should be applied (e.g. calculating market shares and demand elasticities) 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. 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 prof. Giovanni Longo (Università degli Studi di Trieste) for sharing the data, the analysis of the university population and the outcome of the assignment model developed within the framework of the UniMob project - the Mobility Management project for the University of Trieste. The work presented in this paper has greatly benefited from his comments.

7. References

Atasoy, B., Glerum, A., Bierlaire, M. (2011) Mode choice with attitudinal latent class: a swiss case-study, Proceedings of the Second International Choice Modeling Conference, Leeds, UK, July 2011.

Bates, J., Polak, J., Jones, P., Cook, A. (2001) The valuation of reliability for personal travel, *Transportation Research Part E* **37**, 191–229.

Ben-Akiva, M., Lerman, S. R. (1985) *Discrete choice analysis: theory and application to travel demand*, MIT Press.

Ben-Akiva, M., McFadden, D., Garling, T., Gopinath, D., Walker, J., Bolduc, D., Boersch-Supan, A., Delquie, P., Larichev, O., Morikawa, T., Polydoropoulou, A., Rao, V. (1999) Extended framework for modeling choice behavior, *Marketing Letters* **10** (3), 187–203.

Ben-Akiva, M., McFadden, D., Train, K., Walker, J., Bhat, C., Bierlaire, M., Bolduc, D., Boersch-Supan, A., Brownstone, D., Bunch, D., Daly, A., de Palma, A., Gopinath, D., Karlstrom, A., Munizaga, M. A. (2002) Hybrid choice models: Progress and challenges, *Marketing Letters* **13** (3), 163–175.

Bhat, C. (1994) Imputing a continuous income variable from grouped and missing income observations, *Economics Letters* **46**, 311-319.

Bierlaire, M., Fetiarison, M. (2009) Estimation of discrete choice models: extending BIOGEME, Proceedings of the Swiss Transport Research Conference, Ascona, Switzerland, September 2009.

Bolduc, D., Ben-Akiva, M., Walker, J., Michaud, A. (2005) Hybrid choice models with Logit kernel: Applicability to large scale models. In: Lee-Gosselin, M., Doherty, S. (Eds.), *Integrated Land-Use and Transportation Models: Behavioural Foundations*, Elsevier, Oxford, 275-302.

Brownstone, D., Steinmetz, S. (2005) Estimating commuters Value of Time with noisy data. *Transportation Research Part B* **39** (10), 865-889.

Catalano, M., Lo Casto, B., Migliore, M. (2008) Car sharing demand estimation and urban transport demand modelling using stated preference techniques. *European Transport\Trasporti Europei* **40**, 33-50.

Cherchi, E. (2003) *Il Valore del Tempo nella Valutazione dei Sistemi di Trasporto,* Franco Angeli, Milano.

Daly, A., Zachary, S., (1977) The Effect of Free Public Transport on The Journey to Work, Transport and Road Research Laboratory Report SR388.

de Jong, G., Gunn, H. F. (2001) Recent Evidence on Car cost and Time Elasticities of Travel Demand in Europe, *Journal of Transport Economics and Policy* **35** (2), 137-160.

Fiorello, D., Pasti, G. (2003) *Il valore del tempo di viaggio – Guida teorica e applicativa*. Quaderno RT n.5, TRT Trasporti e Territorio, Milano.

Glerum, A., Atasoy, B., Monticone, A., Bierlaire, M. (2011) Adjectives qualifying individuals' perceptions impacting on transport mode preferences, Proceedings of the Second International Choice Modeling Conference, Leeds, UK, July 2011.

Hess, S., Sanko, N., Dumont, J., Daly, A. (2013) A latent variable approach to dealing with missing or inaccurately measured variables: the case of income, Proceedings of the Third International Choice Modelling Conference, July 2013.

Highway Capacity Manual (2010) Transportation Research Board, Washington, D.C.

Horn, J. (1965) A rationale and test for the number of factors in factor analysis. *Psychometrika*, **30** (2), 179-185.

Hornik, J. (1992) Time estimation and orientation mediated by transient mood, *Journal of Socio-Economics* **21** (3), 209–227.

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

Progetto UniMob (2009 -2010) Analisi della popolazione universitaria, Università degli studi di Trieste.

Progetto UniMob (2009 -2010) La domanda di mobilità verso l'Università – Indagine quantitativa, Università degli studi di Trieste.

Progetto UniMob (2009 -2010) Valutazione del sistema di trasporto pubblico, Università degli studi di Trieste.

Rubin, D. (1987) *Multiple Imputation for Nonresponse in Surveys,* Wiley, New York.

Revelle, W. (2012) psych: Procedures for Personality and Psychological Research, R package version 1.2.8, North-Western University, Evanston.

Rietveld, P. (2001) Rounding of Arrival and Departure Times in Travel Surveys, Tinbergen Institute Discussion Paper, TI 2001-110/3.

Rotaris, L., Danielis, R., Rosato, P. (2012) Value of travel time for university students: a revealed / stated preference analysis, *Journal of Environmental Economics and Policy*, 1-21.

Schüssler, N., Axhausen, K. (2011) Psychometric scales for risk propensity, environmentalism and variety seeking, Proceedings of the 9th International Conference on Survey Methods in Transport, November 2011.

Vij, A., Walker, J. (2012) Hybrid choice models: holy grail... or not? Proceedings of the 13th International Conference on Travel Behaviour Research, Toronto, Canada, July 2012.

Walker, J. (2001) *Extended discrete choice models: integrated framework, flexible error structures, and latent variables*, Ph.D. Thesis, Massachusetts Institute of Technology.

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

Walker, J., Li, J., Srinivasan, S., Bolduc, D. (2010) Travel demand models in the developing world: correcting for measurement errors, *Transportation Letters* **2**, 231-243.