Latent variables and route choice behavior

Latent variables and route choice behavior

Carlo Giacomo PRATO *

Department of Transport

Technical University of Denmark

Bygningstorvet 116 Vest, 2800 Kgs. Lyngby, Denmark

Shlomo BEKHOR

Faculty of Civil and Environmental Engineering

Technion – Israel Institute of Technology

Technion Campus, 32000 Haifa, Israel

Cristina PRONELLO

LET : Laboratoire d'Economie des Transports

Université Lumière Lyon 2

14 Avenue Berthelot, 69007 Lyon, France

Dipartimento di Idraulica Trasporti e Infrastrutture Civili

Politecnico di Torino

Corso Duca degli Abruzzi 24, 10142 Torino (TO), Italy

** corresponding author*

Telephone: +45.45256595

Fax: +45.45936533

Email: cgp@transport.dtu.dk

Abstract

In the last decade, a broad array of disciplines has shown a general interest in enhancing discrete choice models by considering the incorporation of psychological factors affecting decision making. This paper provides insight into the comprehension of the determinants of route choice behavior by proposing and estimating a hybrid model that integrates latent variable and route choice models. Data contain information about latent variable indicators and chosen routes of travelers driving regularly from home to work in an urban network. Choice sets include alternative routes generated with a branch and bound algorithm. A hybrid model consists of measurement equations, which relate latent variables to measurement indicators and utilities to choice indicators, and structural equations, which link travelers' observable characteristics to latent variables and explanatory variables to utilities. Estimation results illustrate that considering latent variables (i.e., memory, habit, familiarity, spatial ability, time saving skills) alongside traditional variables (e.g., travel time, distance, congestion level) enriches the comprehension of route choice behavior.

Keywords: Route choice behavior; Latent variables; Hybrid model; Measurement and structural equations; Path size correction logit.

1 Introduction

As the core of traffic assignment and simulation procedures, route choice models allow predicting traffic conditions and forecasting travelers' reactions under future hypothetical scenarios. As the representation of individual behavior, route choice models allow understanding travelers' choices on transportation networks.

The literature in route choice modeling has focused mainly on addressing the "core of traffic assignment" perspective by developing enhanced path generation techniques and discrete choice models. In the first direction, several solutions to the path enumeration problem have been proposed: variations of shortest path algorithms (e.g., Akgün et al. 2000; Hunt and Kornhauser 1997; Lombard and Church 1993; Van der Zijpp and Fiorenzo-Catalano 2005), minimization of generalized cost functions (Ben-Akiva et al. 1984), application of heuristic rules (e.g., Azevedo et al. 1993; De la Barra et al. 1993), single and doubly stochastic simulation approaches (e.g., Bekhor et al. 2006; Bovy and Fiorenzo-Catalano 2007), consideration of logical and behavioral constraints within a branch and bound algorithm (Prato and Bekhor 2006), implementation of a biased random walk algorithm (Frejinger et al. 2009), and combination of breadth first search with network reduction (Schuessler et al. 2010). In the second direction, several solutions to the problem of representing the correlation structure across alternatives have been offered: representation within the deterministic part of the utility function by adding either correction factors (Cascetta et al. 1996) or path size measures (Ben-Akiva and Bierlaire 1999; Bovy et al. 2008), and representation within the stochastic part of the utility function by either relating model parameters to the network topology (Bekhor and Prashker 2001; Prashker and Bekhor 1998) or assuming proportionality between path utility covariance and overlap lengths (Bekhor et al. 2002; Frejinger and Bierlaire 2007; Yai et al. 1997).

The literature in route choice modeling has focused also on addressing the "representation of individual behavior" perspective by presenting route choice models from revealed preference data (e.g., Bekhor et al. 2006; Frejinger and Bierlaire 2007; Hoogendoorn-Lanser 2005; Li et al. 2005; Menghini et al. 2010; Nielsen 2004; Prato 2005; Prato and Bekhor 2006; Ramming 2002; Rich et al. 2007; Wolf et al. 2004). These studies mainly concentrate on the analysis of different applications of path generation techniques and discrete choice models, rather than on the investigation of determinants of individual behavior other than travel times and costs. The only exception is the analysis of the relation between network knowledge and socio-economic factors of travelers with a Multiple Indicator-Multiple Cause model (Ramming 2002), even though without the inclusion of latent variables within the estimated route choice models.

In the last decade, a broad array of disciplines (e.g., psychology, economics, marketing, transportation engineering) has shown a general interest in enhancing discrete choice models by considering the incorporation of psychological factors affecting decision making (Ben-Akiva et al. 2002). A gap still exists between economic modelers, who develop practical models of decision making, and behavioral scientists, who concentrate on the comprehension of agent behavior (Kahneman 2002). In order to bridge this gap, latent constructs need to be incorporated in economic models of decision making (McFadden 2001).

This paper addresses the “representation of individual behavior” perspective and answers the call for incorporating latent constructs in discrete choice models by providing insight into route choice behavior with a hybrid model that integrates latent variable and route choice models. Latent constructs (i.e., memory, habit, familiarity, spatial ability, time saving skills) enter the utility function alongside traditional variables (e.g., travel time, distance) to enrich the comprehension of travelers’ behavior on urban networks.

Behavioral determinants other than travel times and costs have been considered when investigating route diversion, consistency and pre-planning. Madanat et al. (1995) identified the importance of attitudes toward route diversion and perceptions of information reliability on route change following traffic accidents. Abdel-Aty et al. (1995) showed the significant influence of travel time, information reliability and roadway characteristics on route choice between two alternatives. Polydoropoulou et al. (1995) illustrated that a reliable and frequently updated traffic information system primarily affects en-route diversion. Abdel-Aty and Huang (2004) expressed the relevance of travel direction, trip frequency, age and residency on route choices. Bogers et al. (2005) constructed a simulation experiment to explore the influence of information, learning and habit on choices between two routes. Parkany et al. (2006) explained that attitudinal indicators influence consistency and diversion for both stated and revealed preferences of drivers. Ben-Elia et al. (2008) demonstrated that information and personal experience lead to choices between two alternative routes that are different with respect to choices in the same context without any knowledge about the two alternatives. Papinski et al. (2009) examined spatial or temporal deviations between observed and pre-planned routes. While not exhaustive, this list of studies suggests that although recognized as important, latent variables were considered to represent route diversion and planning rather than route choice behavior modeling.

This paper proposes a hybrid model while accounting for spatial abilities and behavioral patterns alongside observable variables, considering several alternatives in a real urban network rather than binary choices in a synthetic experiment, and adopting the framework thoroughly described by Walker (2001) rather than incorporating indicators in utility functions.

Data contain information about travelers who move regularly from home to work in an urban network and participated in a web-based survey. The first part of the survey consisted of four sections of questions: classification of the respondent, investigation of spatial abilities connected to transportation tasks, exploration of spatial abilities not related to transportation tasks, and inquiry of driving preferences. The second part of the survey consisted of the collection of routes considered by the survey participants to drive from home to work.

Route choice sets for modeling purposes contain alternative routes generated with a variation of the branch and bound algorithm (Prato and Bekhor 2006). The proposed variation accounts for the notion that travelers develop their network knowledge by following a transition from landmark recognition to path definition (see Freundschuh 1992; Gale et al. 1990; Garling and Golledge 2000; Golledge and Garling 2003), thus the definition of path similarity shifts from the physical sharing of a number of links to the physical sharing of a number of anchor points through which travelers define their routes.

The hybrid model consists of measurement and structural equations. Measurement equations relate latent variables to measurement indicators and utilities to choice indicators. Structural equations relate travelers’ characteristics to latent variables and observable route attributes and unobservable latent variables to utilities. The latent variable model assumes that indicators are independent on the basis of results from exploratory factor analysis (Prato et al. 2005). The choice model assumes a Path Size Correction Logit formulation (Bovy et al. 2008), since this model allows to account for similarities among alternatives while maintaining the simple Logit structure. The model is estimated through the maximization of a likelihood function that is the integral of the choice model over the distribution of the latent variables.

The remainder of the paper is structured as follows. Section 2 introduces data collection and survey participants. Section 3 describes the choice set generation technique implemented in this study. Section 4 illustrates the structural and the measurement equations composing the hybrid model. Section 5 presents the estimation results and section 6 summarizes major findings of the route choice case study.

2 Data

2.1 Survey design

The data collection process consisted of a web-based survey administered to faculty and staff members of Politecnico di Torino in Italy.

Survey design aimed at being comprehensible. The use of simple and easily understandable language allowed reducing problems related to personal interpretation of the questions. The limitation of the number of questions and the division of the survey in four parts allowed containing the time of survey completion and thus avoiding possible fatigue issue.

The first section included questions about travelers' characteristics such as gender, age, composition of the family, type of employment, level of education and place of residence. The second section investigated spatial abilities involving travel and focused on route learning techniques, perception of travel time for different trip purposes, capacity of memorizing routes under different conditions, and tendency to repeat the same itinerary in different environments. The third section explored spatial abilities not involving travel and concentrated on use of modern search technologies, ability in navigating in different environments, capacity in different memory tasks, and behavior during usual and occasional shopping. The fourth and last section searched for information about knowledge of the city network, capability of estimating distance and time, preferences towards landmark use, highly scenic itineraries and traffic lights avoidance, and preferences towards diversions related to accidents, works or suggestions on the way. Classification questions were formulated in closed form, and latent variable indicators were expressed as Likert-type items of seven points.

A route choice survey accompanied the latent variable survey. Given the purpose of comprehending individual behavior in urban networks, only the urban part of the routes was collected. Initially, each respondent recognized the origin of the trip by individuating on the map either the house location (if resident in Torino) or the access point to the urban network (if resident outside Torino). Then, each respondent identified the common destination by spotting the location of the Politecnico. Last, each respondent indicated the considered routes from home to work by annotating sequences of junctions that were coded on the city map and sending them through a web form. Figure 1 represents an example of coded junctions in proximity of the common destination.

"Insert Figure 1 about here"

Answers to the route choice part of the survey collected information about the routes considered by the respondents. The network for the city of Torino consists of 23 districts, 92 zones, 417 nodes, and 1427 links, and covers an area containing roughly 900,000 inhabitants within the city's limits. The network comprises main roads that cross the town from north to south and from east to west, main arterials that connect different districts of the city, minor arterials that connect points within the same district, and some local streets.

Web-design matched the structure of the survey by preparing an Active Server Page (ASP) page for each section, in which closed-form items presented the text of the question followed by the available alternative answers and latent variable items showed the text of the question followed by a graphic representation of the Likert-scale reporting the semantic meaning of the two extremes (e.g., "difficult…easy"). Automatic recording of identifying session variables and typing actions of the respondents allowed the seamless collection of the answers through the ASP pages. Further details about the web-based survey are presented by Prato et al. (2005).

2.2 Survey participants

Survey participants in the sample for model estimation completed the web-based survey in both the latent variable and the route choice parts. The sample for model estimation consists of 236 individuals indicating a total of 575 routes from home to work, as some respondent provided more than one chosen route in the second part of the survey. Observed routes average 4.8 kilometers in length with a standard deviation of 2.0 kilometers, and 15.4 minutes in time with a standard deviation of 5.9 minutes.

Answers to the first section of the survey provided information about travelers' characteristics that are summarized in table 1. It should be noted that the sample includes mainly males, most likely because of a prevalence of male population among faculty and staff members in the Politecnico di Torino, and graduated respondents, most probably because the participation in the survey of faculty members skewed the sample from the education level perspective.

"Insert Table 1 about here"

Answers to the second through the fourth part of the survey provided information about latent variable indicators. Measures of internal consistency and sampling adequacy (Prato et al. 2005) showed the suitability of 28 indicators for modeling purposes, according to their high internal consistency throughout the entire latent variable survey (Cronbach's Alpha = 0.76), and their high adequacy at the item level (Kaiser-Meyer-Olkin > 0.7). Moreover, exploratory factor analysis helped individuating latent variables likely affecting route choice behavior (Prato et al. 2005) and showed each indicator having high factor loading on only one latent factor. Accordingly, the latent variables and the related indicators are presented in table 2.

"Insert Table 2 about here"

3 Choice set generation

Modeling route choice behavior usually consists of the individuation of available alternative routes and the calculation of the probability of choosing a certain route from the generated choice set.

While the web-based survey collected the chosen routes, the implementation of the branch and bound algorithm (Prato and Bekhor 2006) allowed generating routes alternative to the chosen ones reported by survey participants. The algorithm explicitly constructs a connection tree between origin and destination of each registered trip by processing sequences of links according to a branching rule that accounts for behavioral and logical constraints. Each sequence of links reaching the destination while satisfying all the constraints enters the choice set of each observation as a feasible solution to the path enumeration problem.

The following logical and behavioral constraints were considered for path generation purposes:

- A *directional constraint* excludes from consideration paths containing links that take the driver farther from the destination and closer to the origin, with a tolerance equal to 10%.
- A *temporal constraint* rejects paths that travelers would consider unrealistic since their travel time is excessively higher than the shortest path, with a tolerance equal to 50% travel time in excess.
- A *loop constraint* discards path segments that travelers would not consider because they constitute a detour larger than an acceptable value, with a tolerance equal to 10% extra time for detours.
- A *similarity constraint* removes highly overlapping paths that travelers would not consider as separate alternatives. Specifically, paths are considered similar when sharing more than 3 common landmarks that are defined as the intersections between the major arterials according to the city road hierarchy.
- A *movement constraint* eliminates unrealistic path segments causing delay and apprehension in drivers approaching the junction. Specifically, a movement threshold limits to 4 the number of left turns in signalized intersections since traffic light regulation in Torino does not reserve green time for left turns.

Note that the definition of the constraints differs slightly from the original formulation proposed by Prato and Bekhor (2006), as this variation of the algorithm accounts for the notion that travelers develop their network knowledge by following a transition from landmark recognition to path definition (see Freundschuh 1992; Gale et al. 1990; Garling and Golledge 2000; Golledge and

Garling 2003). According to this notion, travelers navigate through landmarks and might consider as similar paths that share the same sequences of landmarks. Hence, the similarity constraint considers that routes are alike not because of their physical sharing of a number of links, but because of their physical sharing of a number of anchor points through which travelers define them.

The branch and bound algorithm produced a set of alternatives for each observation by processing the origin-destination pair of each of the 575 observed routes recorded in the survey. The number of generated alternatives varies between 2 and 19 with a median value of 11 alternative routes per observation. The comparison of the generated choice sets with the observed routes reveals that the coverage (see Ramming 2002) is 85.4% with a 100% overlap threshold and 91.3% with an 80% overlap threshold. Associating these values with the consideration that all observed routes overlap at least 64.5% with the generated routes shows high realism of the implemented path generation technique with respect to the observed behavior. It should be noted that the observed routes not reproduced at the 80% overlap threshold were added to the generated choice set.

As the impact of choice set size and composition on model estimates has recently received attention (e.g., Prato and Bekhor 2007; Bliemer and Bovy 2008), alternative choice set generation techniques were implemented in order to perform a sensitivity analysis of model estimates with respect to choice set composition. The random walk algorithm (Frejinger et al. 2009) was implemented with both parameters of the Kumaraswamy distribution equal to one for 50 iterations. Choice sets contain between 3 and 35 alternatives with a median value of 17, and the coverage is 78.4% with a 100% overlap threshold and 87.3% with an 80% overlap threshold. The breadth first search on link elimination (BFS-LE) algorithm (Schuessler et al. 2010) was applied with shuffling of the sub-network list. Choice sets include between 2 and 21 routes with a median value of 11, and the coverage is 80.7% with a 100% overlap threshold and 90.6% with an 80% overlap threshold.

4. Hybrid model

4.1 Model formulation

The hybrid model framework integrates a latent variable model and a route choice model. The latent variable model consists of structural equations, which describe the latent variables as a function of observable individual characteristics, and measurement equations, which relate the unobservable latent variables to observable indicators. The route choice model consists of structural equations, which link observable and latent variables to the route utilities, and measurement equations, which express the choice as a function of the unobservable utilities. Figure 2 represents the hybrid model framework inspired from the original framework proposed by Walker (2001).

"Insert Figure 2 about here"

The structural equations of the latent variable model express the distribution of the latent variables (Walker 2001):

$$X_n^* = g_1\left(S_n;\gamma\right) + \omega_n \quad \text{and} \quad \omega_n \sim D\left(0,\Sigma_\omega\right) \tag{1}$$

where X^*_n is a vector of latent variables, S_n is a vector of characteristics of individual n, ω_n is a vector of error terms following distribution D with covariance matrix Σ_ω, and γ is a matrix of parameters to be estimated.

The structural equations of the choice model express the distribution of the utilities (Walker 2001):

$$U_n = V\left(Z_n, X_n^*;\beta\right) + \varepsilon_n \quad \text{and} \quad \varepsilon_n \sim D\left(0,\Sigma_\varepsilon\right) \tag{2}$$

where U_n is a vector of utilities of alternative routes, Z_n is a vector of attributes of alternative routes, ε_n is a vector of error terms following distribution D with covariance matrix Σ_ε, and β is a vector of parameters to be estimated.

The measurement equations of the latent variable model express the distribution of the indicators (Walker 2001):

$$I_n = g_2\left(X_n^*;\alpha\right) + \upsilon_n \quad \text{and} \quad \upsilon_n \sim D\left(0,\Sigma_\upsilon\right) \tag{3}$$

where I_n is a vector of indicators, υ_{rn} is a vector of error terms following distribution D with covariance matrix Σ_υ, and α is a vector of parameters to be estimated.

The measurement equations of the choice model express the choice as a function of the utilities (Walker 2001):

$$y_{in} = \begin{cases} 1 & \text{if } U_{in} \geq U_{jn} \ \forall j \neq i \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where y_{in} is the indicator of choosing route i over alternative routes j, and U_{in} is the utility of route i.

The estimation of the hybrid model is performed by maximum simulated likelihood. If the latent variables were not present, the choice probability $P(y_n \mid Z_n, \beta, \Sigma_\varepsilon)$ of selecting the observed routes would be sufficient to write the likelihood function. As the latent variables are present in the hybrid model, the choice probability should be expressed as $P(y_n \mid X^*_n, Z_n, \beta, \Sigma_\omega, \Sigma_\varepsilon)$, but since latent variables are not actually observed, the choice probability is obtained by integrating over the distribution of the latent variables:

$$P\left(y_n \middle| Z_n, S_n, \beta, \gamma, \Sigma_\varepsilon, \Sigma_\omega\right) = \int_{X_n^*} P\left(y_n \middle| X_n^*, Z_n, \beta, \Sigma_\varepsilon\right) f_1\left(X_n^* \middle| S_n, \gamma, \Sigma_\omega\right) dX_n^* \tag{5}$$

where $f_1(X^*_n \mid S_n, \gamma, \Sigma_\omega)$ is the density function of the latent variables.

Since indicators are observed, the joint probability of observing choice and latent variable indicators is written as:

$$\begin{aligned} &P\left(y_n, I_n \middle| Z_n, S_n, \beta, \alpha, \gamma, \Sigma_\varepsilon, \Sigma_\upsilon, \Sigma_\omega\right) = \\ &= \int_{X_n^*} P\left(y_n \middle| X_n^*, Z_n, \beta, \Sigma_\varepsilon\right) f_2\left(I_n \middle| X_n^*, \alpha, \Sigma_\upsilon\right) f_1\left(X_n^* \middle| S_n, \gamma, \Sigma_\omega\right) dX_n^* \end{aligned} \tag{6}$$

where $f_2\ (I_n \mid X^*_n, \alpha, \Sigma_\upsilon)$ is the density function of the indicators.

In this hybrid model, the functional form of the route choice model is a Path Size Correlation Logit (Bovy et al. 2008) that allows accommodating the correlation across alternative routes while maintaining the simple Logit structure:

$$P\left(y_{in} \middle| X_n^*, Z_n, \beta, \Sigma_\varepsilon\right) = \frac{\exp\left(Z_{in}\beta_{obs} + X_{in}^*\beta_{lat} + PSC_i\beta_{PSC}\right)}{\sum_j \exp\left(Z_{jn}\beta_{obs} + X_{jn}^*\beta_{lat} + PSC_j\beta_{PSC}\right)} \tag{7}$$

where Z_{in} is a vector of attributes of route i, X^*_{in} is a vector of latent variables associated to route i, PSC_i is the path size correction of route i, β_{PSC} is a parameter related to the path size correction, β_{obs} is a vector of parameters related to the observable route attributes, and β_{lat} is a vector of parameters related to the latent variables. The path size correction is calculated as (Bovy et al. 2008):

$$PSC_i = -\sum_{a \in \Gamma_i} \left(\frac{D_a}{D_i} \ln \sum_j \delta_{aj} \right) \tag{8}$$

where D_i is the length of route i, D_a is the length of link a within the set of links Γ_i, and δ_{aj} is the link-path incidence dummy equal to one if route j uses links a and zero otherwise.

In this hybrid model, the densities of the latent variables and the indicators are expressed as follows:

$$f_1\left(X_n^* \middle| S_n, \gamma, \sigma_\omega \right) = \prod_{l=1}^{L} \frac{1}{\sigma_{\omega_l}} \phi \left(\frac{X_{ln}^* - S_{ln}\gamma_l}{\sigma_{\omega_l}} \right) \tag{9}$$

$$f_2\left(I_n \middle| X_n^*, \alpha, \sigma_\upsilon \right) = \prod_{r=1}^{R} \frac{1}{\sigma_{\upsilon_r}} \phi \left(\frac{I_{rn} - X_{ln}^*\alpha_r}{\sigma_{\upsilon_r}} \right) \tag{10}$$

where S_{ln} is a vector of individual characteristics related to one of L latent variables, I_{rn} is one of R indicators, $\sigma_{\omega l}$ and σ_{υ_r} are variances of error terms in vectors ω and υ, α_r and γ_l are parameters respectively related to indicators and latent variables, and Φ is the standard normal density function.

Given the expectation form, the choice probability may be replaced by an empirical mean that simulates the L-dimensional integral:

$$\tilde{P}\left(y_n, I_n \middle| Z_n, S_n, \beta, \alpha, \gamma, \Sigma_\varepsilon, \Sigma_\upsilon, \Sigma_\omega \right) = \\ \frac{1}{H} \sum_{h=1}^{H} \frac{\exp\left(Z_{in}\beta_{obs} + X_{in}^{*h}\beta_{lat} + PSC_i\beta_{PSC} \right)}{\sum_j \exp\left(Z_{jn}\beta_{obs} + X_{jn}^{*h}\beta_{lat} + PSC_j\beta_{PSC} \right)} \prod_{r=1}^{R} \frac{1}{\sigma_{\upsilon_r}} \phi \left(\frac{I_{rn} - X_{ln}^*\alpha_r}{\sigma_{\upsilon_r}} \right) \tag{11}$$

where H is the number of draws and X_{jn}^{*h} is a random draw of the latent variable l that is calculated as:

$$X_{ln}^{*d} = S_{ln}\gamma_l + \omega_{ln}^d = S_{ln}\gamma_l + \sigma_{\omega_l}\tilde{\omega}_{ln} \quad \text{where} \quad \tilde{\omega}_{ln} \approx N(0,1) \tag{12}$$

The objective function becomes:

$$\max_{\alpha,\beta,\gamma} \sum_{n=1}^{N} \ln \tilde{P}\left(y_n, I_n \middle| Z_n, S_n, \beta, \alpha, \gamma, \Sigma_\varepsilon, \Sigma_\upsilon, \Sigma_\omega\right) \tag{13}$$

The maximization of the likelihood function is performed simultaneously by simulating the integration of the choice model over the distribution of the fitted latent variables with code written in Gauss matrix language. Following the literature about integral simulation (e.g., Bhat 2003; Train 2003) and about hybrid model estimation (e.g., Walker 2001; Bolduc et al. 2008), 1000 Halton draws are used for the simulation of the *L*-dimensional integral.

4.2 Model specification

Measurement equations of the latent variable model associate the latent variables to the indicators according to the correspondence in table 2, and as an example the first of the 28 equations is presented:

$$MEMROUTE_n = MEM_n \alpha_1 + \upsilon_{1n} \tag{14}$$

Structural equations of the latent variable model associate the latent variables to the individual characteristics, and after statistical significance tests for the parameters within γ, some parameters were constrained to zero and the five structural equations of the latent variable model are written as:

$$\begin{aligned} MEM_n = MALE_n\gamma_{1,1} + AGEL35_n\gamma_{1,2} + AGEM55_n\gamma_{1,3} + EDUC_n\gamma_{1,4} + \\ CHILDREN_n\gamma_{1,6} + STOP_n\gamma_{1,7} + CONST_n\gamma_{1,9} + \omega_{1n} \end{aligned} \tag{15}$$

$$\begin{aligned} HAB_n = MALE_n\gamma_{2,1} + AGEL35_n\gamma_{2,2} + AGEM55_n\gamma_{2,3} + EDUC_n\gamma_{2,4} + \\ SINGLE_n\gamma_{2,5} + STOP_n\gamma_{2,7} + CONST_n\gamma_{2,9} + \omega_{2n} \end{aligned} \tag{16}$$

$$\begin{aligned} FAM_n = MALE_n\gamma_{3,1} + CHILDREN_n\gamma_{3,6} + STOP_n\gamma_{3,7} + \\ RESCITY_n\gamma_{3,8} + CONST_n\gamma_{3,9} + \omega_{3n} \end{aligned} \tag{17}$$

$$\begin{aligned} SPAB_n = MALE_n\gamma_{4,1} + EDUC_n\gamma_{4,4} + CHILDREN_n\gamma_{4,6} + \\ STOP_n\gamma_{4,7} + RESCITY_n\gamma_{4,8} + CONST_n\gamma_{4,9} + \omega_{4n} \end{aligned} \tag{18}$$

$$\begin{aligned} TSAV_n = MALE_n\gamma_{5,1} + AGEL35_n\gamma_{5,2} + AGEM55_n\gamma_{5,3} + EDUC_n\gamma_{5,4} + \\ SINGLE_n\gamma_{5,5} + STOP_n\gamma_{5,7} + RESCITY_n\gamma_{5,8} + CONST_n\gamma_{5,9} + \omega_{5n} \end{aligned} \tag{19}$$

where $MALE_n$ indicates the gender (equal to 1 if male, 0 if female), $AGEL35_n$ and $AGEL55_n$ refer to the age (less than 35 or more than 55 years old, respectively), $EDUC_n$ denotes the education level (equal to 1 if at least M.Sc., 0 otherwise), $SINGLE_n$ and $CHILDREN_n$ represent the family status (single or married with children, respectively), $RESCITY_n$ indicates the residence location (equal to 1 if within the city, 0 otherwise), and $STOP_n$ refers to stops along the commute trip (equal to 1 if usual, 0 otherwise).

Structural equations of the choice model associate route utilities with route attributes and latent variables as perceived by individual n:

$$\begin{aligned} V_{jn} = {} & DIST_{jn}\beta_1 + TIME_{jn}\beta_2 + DELPC_{jn}\beta_3 + TMRPC_{jn}\beta_4 + TURNS_{jn}\beta_5 + \\ & + PSC_{jn}\beta_6 + MEM_n DIST_{jn}\beta_7 + MEM_n DELPC_{jn}\beta_8 + HAB_n DIST_{jn}\beta_9 + \\ & + HAB_n TMRPC_{jn}\beta_{10} + HAB_n TURNS_{jn}\beta_{11} + FAM_n DELPC_{jn}\beta_{12} + \\ & + FAM_n TMRPC_{jn}\beta_{13} + FAM_n TURNS_{jn}\beta_{14} + SPAB_n TMRPC_{jn}\beta_{15} + \\ & + SPAB_n TURNS_{jn}\beta_{16} + TSAV_n TIME_{jn}\beta_{17} + TSAV_n DELPC_{jn}\beta_{18} \end{aligned} \tag{20}$$

where $DIST_{jn}$ is the distance, $TIME_{jn}$ is the travel time, $DELPC_{jn}$ is the percentage of delay, $TMRPC_{jn}$ is the percentage of time on major roads, $TURNS_{jn}$ is the number of turns and PSC_{jn} is the path size correction factor of the alternative route j within the choice set of individual n. The values of the latent variables MEM_n, HAB_n, FAM_n, $SPAB_n$ and $TSAV_n$ for each respondent are associated to each route recorded through interaction terms with the route attributes. A systematic process of considering every possible interaction term between latent variables and route attributes and examining the significance of the estimated parameters led to the significant interaction terms in equation (20).

The measurement equations of the choice model individuate the chosen routes within the sets of alternative routes.

5 Model results

5.1 Latent variable model

Estimates of the measurement equations are presented in table 3, where 5 parameters are constrained to one for identification purposes (see Walker 2001) and estimates of the 28 standard deviations σ_{υ_r} are not reported.

"Insert Table 3 about here"

As the model measures the effects of the latent variables on each indicator, some considerations are drawn from the results. As expected, the latent variable MEM has a positive correlation with all memorizing tasks and especially with transportation related tasks such as remembering a route just learned, a route traveled as a passenger or a parking location. The latent variable HAB is positively linked to the habit of driving through the same route and the recurrence of shopping in the same places, and is negatively linked to the tendency to modify itinerary as a consequence of either traffic congestion or received information. The correlation of the latent variable FAM is positive at a large extent with the ability of describing routes usually taken and evaluating travel time of any route, is positive at a smaller extent with the capability of navigating at home in the dark, and is negative with the tendency of using main roads for navigation across the city. The correlation of the latent variable SPAB is positive with the ability of evaluating distances on a map, at a lesser extent positive with using maps and navigating through landmarks, and is negative with the preference for scenic roads. The latent variable TSAV is positively linked at a larger extent to the search for shortcuts and the preference for routes without traffic lights, and at a lesser extent to the tendency of properly estimating times and distances.

Estimates of the structural equations are presented in table 4 and estimates of the 5 standard deviations σ_{ω_l} are not reported. It should be noted that covariances of the latent variables are constrained to zero, after initial unconstrained estimation of the model verified that estimates of covariances are not significantly different from zero. The orthogonality of the latent variables confirms analogous findings by Prato et al. (2005).

"Insert Table 4 about here"

As the model links travelers' characteristics with the latent variables, some considerations are elicited from the results. Being a male is related to higher mnemonic capability, higher level of familiarity with the environment, better spatial abilities and superior time saving skills. Younger respondents seem expectedly related to having both better memory and time saving skills and appear understandably connected to a lower tendency to follow routine behavior, while older respondents seem logically related to opposite tendencies. Having obtained a degree seems understandably correlated to higher abilities in terms of memory, spatial orientation ability and time saving skills, but also to higher propensity to be habit-bound. Family composition shows also association to the latent variables, as being single seems predictably related to less habitual behavior and not expectedly connected to lower time saving skills, while having children appears to relate to higher mnemonic ability and familiarity with the choice environment. Having stops on the way to work is linked positively to habit and negatively to spatial abilities, but is also less expectedly related positively to time saving skills and negatively to memory and familiarity of the environment. Last, being resident in the city seems logically associated positively with higher familiarity with the environment, routine behavior and ability in saving time, and negatively with spatial abilities.

5.2 Route choice model

Estimates of the route choice model are presented in table 5, alongside the estimates of a stand-alone route choice model without latent variables.

"Insert Table 5 about here"

Notably, the inclusion of the latent variables identified by the structural equations improves the goodness-of-fit of the hybrid model with respect to the stand-alone choice model. As the number of parameters of the hybrid model is much larger, an account of prediction performance of the models is given by applying the models to estimation and validation samples. Prediction involved drawing randomly 475 observations for estimation purposes and 100 observations for validation purposes, repeating the procedure 10 times for reducing the effect of the random draws, and computing probabilities of choosing each alternative route in order to calculate the average probability of correctly predicting the choice of each observation. While the application of the PSC-Logit model is straightforward, the application of the hybrid model implies the integration of the choice model over the distribution of the disturbances of the structural equations of the latent variable model (Walker 2001). The hybrid model outperforms the stand-alone route choice model in terms of average probability of correct prediction (32.4% versus 25.0%). It should be noted that the relatively low values must be put into perspective by considering that an average of 11 routes are available for each observation in the validation sample. It should be also noted that the average overlap of the predicted routes (i.e., routes with the highest choice probability in the validation sample) with respect to the chosen routes further confirms that the hybrid model outperforms the stand-alone route choice model (78.1% versus 70.3%)

Parameter estimates of the route attributes suggest that increasing distances and travel times have an expected negative effect on the selection of a route. Logically, the same applies to the percentage of delay that measures the average level of congestion on the route as the ratio of the difference between congested and free flow time with respect to the congested travel time. Also logically, the same concerns the number of turns in accordance with the notion of travelers preferring direct routes. Plausibly, the percentage of time on major roads is positively related to route choices of individuals in accordance to the notion that travelers prefer to navigate through landmarks and in this specific case through major arterials. The sign of the parameter of the path size correction factor is positive, to confirm the desired reduction in the utility of overlapping routes.

Parameter estimates of the interaction terms between latent variables and route attributes suggest that mnemonic, spatial and time saving abilities seem to have a positive correlation with the described preferences of individuals for the route attributes, while habit and familiarity appear to have a negative one. On the one hand in fact, individuals with higher mnemonic capacity seem to look for shorter and less congested paths, better spatial ability is not surprisingly related to a larger use of landmarks and a lower number of turns, and travelers with high time saving skills appear to tend toward faster and less congested alternatives. On the other hand, habit-bound travelers seem not to care about longer distances, lower use of major arterials and higher number of turns in their route choices, while individuals highly familiar with the environment in which they travel appear less bothered by higher congestion levels, lower use of landmarks and increasing turning movements.

Table 6 presents a sensitivity analysis of model estimates with respect to the choice set generation technique. Not only the parameter signs are not different, but also the parameter estimate values are not significantly different when estimating the hybrid model with choice sets generated by either branch and bound (Prato and Bekhor 2006), or random walk (Frejinger et al. 2009), or BFS-LE (Schuessler et al. 2010) algorithms. Even though literature in route choice modeling shows that choice set composition affects model estimates (e.g., Prato and Bekhor 2007; Bliemer and Bovy 2008), the hybrid model seems not to be affected. On the one hand, this finding could be explained by the fact that choice set sizes and compositions are comparable for this specific case-study. On the other hand, this finding could be explained by the fact that latent variables are invariant to the choice set generation technique.

"Insert Table 6 about here"

6 Summary and conclusions

This paper provides insight into route choice behavior by estimating a hybrid latent variable choice model where latent constructs (i.e., memory, habit, familiarity, spatial ability, time saving skills) enter the utility function alongside traditional variables (e.g., travel time, distance, congestion level) to enrich the comprehension of individual behavior on urban networks.

The collection of latent variable indicators, the recording of chosen routes and the generation of choice sets provided the data for modeling purposes. The design of a web-based survey allowed collecting travelers' characteristics, transportation and non-transportation related spatial abilities, behavioral patterns, and 575 routes chosen by the survey respondents to drive from home to work. A modification of the branch and bound algorithm (Prato and Bekhor 2006) accounted for the notion that travelers perceive similarity among paths on the basis not only of the physical sharing of a number of links, but also of the physical sharing of a number of anchor points through which they define their routes. Random walk (Frejinger et al. 2009) and BFS-LE (Schuessler et al. 2010) algorithms allowed performing a sensitivity analysis of hybrid model estimates with respect to the choice set generation technique implemented.

Simultaneous estimation of the hybrid model allowed estimating the parameters of both the latent variable and the route choice models.

Notably, the inclusion of the latent variables improves the goodness-of-fit of the hybrid model over the stand-alone choice model, as shown by both goodness-of-fit measures and prediction performances.

Expectedly, increasing distances, travel times, congestion levels and number of turns have a negative effect on the choice of routes, while higher percentage of time on major roads has a positive effect. These results confirm well known findings about minimization of travel time and distance (e.g., Ramming 2002; Hoogendoorn-Lanser 2005; Prato 2005), minimization of congestion levels (e.g., Prato 2005; Papinski et al. 2009) and maximization of route directness (Raghubir and Krishna 1996; Conroy Dalton 2003; Papinski et al. 2009).

Latent variables provide additional insight into route choice behavior by suggesting that mnemonic, spatial and time saving abilities seem to have a positive correlation with the preferences of individuals with respect to route attributes in the sense that probably individuals with these skills tend to look for better alternatives and to remember to use them. On the other hand, habit and familiarity appear to have a negative correlation with the preferences of individuals for route attributes in the sense that possibly individuals with these characteristics do not tend to search for better alternative routes even if their choice is not optimal.

Probably, being able to search for alternatives that allow saving time and being able to remember several available alternatives may increase the utility of route choices in the sense that individuals tend to look for better alternatives and to remember using them. In fact, greater spatial knowledge is related to greater variation in route (Ramming 2002), as travelers with better spatial abilities might be aware of more routes and look for information to decide among them, or might listen to travel reports and read maps to acquire additional alternatives.

Presumably, having the habit of following the same route and navigating mainly through familiar places may reduce the utility of route choices in the sense that individuals do not tend to search for better alternative routes even if their choice is not optimal in terms of travel time, congestion and number of turns. In fact, commuting route choices are frequent choices that are goal-directed habitual behavior (Aarts and Dijksterhuis 2000) and hence are characterized by automaticity and partial lack of awareness (Verplanken and Aarts1999) that leads travelers not necessarily to minimize distance and travel time.

These results confirm previous findings found in the travel behavior literature. Model estimates suggest that latent variables alter the perception of alternative attributes by travelers (e.g., habit leads to choosing longer routes), as hypothesized in the conceptual framework proposed by Bovy and Stern (1990) to describe the route choice process as dependent not only on route attributes, but also on spatial abilities, behavioral patterns and driving preferences of travelers. Also, the significance of the latent factors generalize to route choice from revealed preferences findings about route choice in simulation experiments, performed for example by Polydoropoulou et al. (1995), who determined the influence of attitudes on route choice diversions, Bogers et al. (2005), who analyzed the effect of learning and habit in a simulation of selection between two routes, and Parkany et al. (2006), who illustrated that attitudes affect consistency and diversion in the choice of paths. Moreover, the relevance of habit and familiarity agrees with the theory that behavior really has core preferences based on habitual behavior and contingent preferences based on context (Fujii and Garling 2003).

Undoubtedly, estimating a hybrid model contributes to the understanding of determinants of individual route choice behavior in urban networks. Findings suggest that individuals generally prefer shorter, faster and less congested routes, but also that their characteristics, their spatial abilities and their behavioral patterns significantly influence their preferences and could even bring them to ignore better alternatives because they are comfortable with their current choices. Further research could concentrate on the simplification of the model specification with a lower number of variables, on the consideration of the concept of landmark similarity in the route choice model and not only in the choice set generation, and on the analysis of the effect of latent variables on the choice set formation process when a joint model of choice set generation and route choice is proposed and estimated.

Acknowledgements

The authors would like to express their gratitude to Kay Axhausen and Michel Bierlaire for commenting on a preliminary version of the paper and suggesting improvements to the model specification. The authors would also like to express their appreciation for the contribution of four anonymous reviewers.

References

Aarts H, Dijksterhuis AP (2000) The automatic activation of goal-directed behaviour: the case of travel habit. J Environ Psychol, 20(1):75-82.

Abdel-Aty MA, Huang Y (2004) Exploratory spatial analysis of expressway ramps and its effects on route choice. J Transp Eng 130(1):104-112

Abdel-Aty MA, Kitamura R, Jovanis PP (1995) Exploring route choice behavior using geographic information system-based alternative routes and hypothetical travel time information input. Transp Res Rec 1493:74-80

Akgün V, Erkut E, Batta R (2000) On finding dissimilar paths. Eur J Oper Res 121(2):232-246

Azevedo JA, Santos Costa MEO, Silvestre Madera JJER, Vieira Martins EQ (1993) An algorithm for the ranking of shortest paths. Eur J Oper Res 69(1):97-106

Bekhor S, Ben-Akiva ME, Ramming S (2002) Adaptation of logit kernel to route choice situation. Transp Res Rec 1805:78-85

Bekhor S, Ben-Akiva ME, Ramming S (2006) Evaluation of choice set generation algorithms for route choice models. Ann Oper Res 144:235-247

Bekhor S, Prashker JN (2001) Stochastic user equilibrium formulation for the generalized nested logit model. Transp Res Rec 1752:84-90

Ben-Akiva ME, Bergman MJ, Daly AJ, Ramaswamy R (1984) Modeling interurban route choice behaviour. In Volmuller J, Hamerslag R (Eds.), Proceedings of the 9th International Symposium on Transportation and Traffic Theory, Utrecht (The Netherlands): VNU Science Press, pp 299-330

Ben-Akiva ME, Bierlaire M (1999) Discrete choice methods and their applications to short term travel decisions. In Hall RW (Ed.), Handbook of transportation science, Dordrecht (The Netherlands): Kluwer, pp 5-34

Ben-Akiva ME, McFadden D, Train KE, Walker JL, Bhat CR, Bierlaire M, Bolduc D, Boersch-Supan A, Brownstone D, Bunch D, Daly A, de Palma A, Gopinath D, Karlstrom A, Munizaga MA (2002) Hybrid choice models: progress and challenges. Mark Lett 13(3):163-175

Ben-Elia E, Erev I, Shiftan Y (2008) The combined effect of information and experience on drivers' route-choice behavior. Transp 35(2):165-177

Bhat CR (2003) Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. Transp Res Part B 37(9):837-855

Bliemer MCJ, Bovy PHL (2008) Impact of route choice set on route choice probabilities. Transp Res Rec 2076:10-19

Bogers EAI, Viti F, Hoogendoorn SP (2005) Joint modeling of advanced travel information service, habit, and learning impacts on route choice by laboratory simulator experiments. Transp Res Rec 1926:189-197

Bolduc D, Boucher N, Alvarez-Daziano R (2008) Hybrid choice modeling of new technologies for car choice in Canada. Transp Res Rec 2082:63-71

Bovy PHL, Bekhor S, Prato CG (2008) The factor of revisited path size: an alternative derivation. Transp Res Rec 2076:132-140

Bovy PHL, Fiorenzo-Catalano S (2007) Stochastic route choice set generation: behavioral and probabilistic foundations. Transportmetrica 3(3):173-189

Bovy PHL, Stern E (1990) Route choice: wayfinding in transport networks. Dordrecht (The Netherlands): Kluwer Academic Publishers

Cascetta E, Nuzzolo A, Russo F, Vitetta A (1996) A modified logit route choice model overcoming path overlapping problems: specification and some calibration results for interurban networks. In: Lesort JB (Ed.), Proceedings of the 13th International Symposium on Transportation and Traffic Theory, Lyon (France): Pergamon Press, pp 697-711

Conroy Dalton R (2003) The secret is to follow your nose: route path selection and angularity. Environ Behav 35(1):107-131.

De La Barra T, Perez B, Anez J (1993) Multidimensional path search and assignment. Proceedings of the 21st PTRC Summer Annual Meeting (pp. 307-319). Manchester: PTRC.

Frejinger E, Bierlaire M (2007) Capturing correlation with subnetworks in route choice models. Transp Res Part B 41(3):363-378

Frejinger E, Bierlaire M, Ben-Akiva ME (2009) Sampling of alternatives for route choice modeling. Transp Res Part B 43(10):984-994

Freundschuh SM (1992) Is there a relationship between spatial cognition and environmental patterns? Lect Notes Comput Sci 639:288-304

Fujii S, Garling T (2003) Application of attitude theory for improved predictive accuracy of stated preference methods in travel demand analysis. Transp Res Part A 37(4):389-402

Gale N, Golledge RG, Pellegrino JW, Doherty S (1990) The acquisition and integration of route knowledge in an unfamiliar neighbourhood. J Environ Psychol 10:3-25

Garling T, Golledge RG (2000) Cognitive mapping and spatial decision making. In Kitchin R, Freundschuh S (Eds.), Cognitive mapping: past, present, and future, London: Routledge, pp 44-65

Golledge RG, Garling T (2003) Spatial behavior in transportation modeling and planning. In Goulias KG (Ed.), Transportation systems planning: methods and applications, New York (USA): CRC Press, pp 3-1 – 3-27

Hoogendoorn-Lanser S (2005) Modelling travel behaviour in multi-modal networks. Ph.D. Thesis, TRAIL Research School, Technical University of Delft, The Netherlands.

Hunt DT, Kornhauser AL (1997) Assigning traffic over essentially-least-cost paths. Transp Res Rec 1556:1-7

Kahneman D (2002) Maps of bounded rationality: a perspective on intuitive judgment and choice. In: Frangsmyr T [Nobel Foundation] (Ed.), Les Prix Nobel: the Nobel Prizes 2002, Stockholm (Sweden): The Nobel Foundation, pp 449-489

Li H, Guensler R, Ogle J (2005) Analysis of morning commute route choice patterns using Global Positioning System-based vehicle activity data. Transp Res Rec 1926:162-170.

Lombard K, Church RL (1993) The gateway shortest path problem: generating alternative routes for a corridor location problem. Geogr Syst 1:25-45

Madanat SM, Yang CY, Yen YM (1995) Analysis of stated route diversion intentions under advanced traveler information systems using latent variable modeling. Transp Res Rec 1485:10-17

McFadden D (2001) Economic Choices. Am Econ Rev 91(3):351-378

Menghini G, Carrasco N, Schuessler N, Axhausen KW (2010). Route choice of cyclists in Zurich. Transp Res Part A 44(9):754-765.

Nielsen OA (2004) Behavioural responses to pricing schemes: description of the Danish AKTA experiment. J Intell Transp Syst 8(4): 233-251.

Papinski D, Scott DM, Doherty ST (2009) Exploring the route choice decision-making process: a comparison of pre-planned and observed routes obtained using person-based GPS. Transp Res Part F 12(4):347-358

Parkany E, Du J, Aultman-Hall L, Gallagher R (2006) Modeling stated and revealed route choice: consideration of consistency, diversion, and attitudinal variables. Transp Res Rec 1985:29-39

Polydoropoulou A, Ben-Akiva ME, Kaysi I (1995) Influence of traffic information on drivers' route choice behavior. Transp Res Rec 1453:56-65

Prashker JN, Bekhor S (1998) Investigation of stochastic network loading procedures. Transp Res Rec 1645:94-102

Prato CG (2005) Latent factors and route choice behaviour. Ph.D. Thesis, Politecnico di Torino, Italy.

Prato CG, Bekhor S (2006) Applying branch & bound technique to route choice set generation. Transp Res Rec 1985:19-28

Prato CG, Bekhor S (2007) Modeling route choice behavior: how relevant is the choice set composition? Transp Res Rec 2003:64-73

Prato CG, Bekhor S, Pronello C (2005) Methodology for exploratory analysis of latent factors influencing drivers' behavior. Transp Res Rec 1926:115-125

Raghubir P, Krishna A (1996) As the crow flies: bias in consumers' map-based distance judgments. J Cons Res 23(1):26-39.

Ramming S (2002) Network knowledge and route choice. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, USA.

Rich J, Mabit S, Nielsen OA (2007) Route choice model for Copenhagen: a data-driven choice set generation approach based on GPS data. Proceedings of the 6th Tristan Conference, Phuket, Thailand.

Schuessler N, Balmer M, Axhausen KW (2010) Route choice sets for very high-resolution data. Proceedings of 89th Annual Meeting of the Transportation Research Board, Washington, DC.

Train K (2003) Discrete choice methods with simulation. Cambridge (UK): Cambridge University Press.

Van der Zijpp NJ, Fiorenzo-Catalano S (2005) Path enumeration by finding the constrained K-shortest paths. Transp Res Part B 39(6):545-563

Verplanken B, Aarts H (1999) Habit, attitudes, and planned behaviour: is habit an empty construct or an interesting case of goal-directed automaticity? Eur Rev Soc Psychol, 10:101-134.
Walker JL (2001) Extended discrete choice models: integrated framework, flexible error structures, and latent variables. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, USA.
Wolf J, Schönfelder S, Samaga U, Oliveira M, Axhausen KW (2004) Eighty weeks of Global Positioning System traces: approaches to enriching trip information. Transp Res Rec 1870:46-54.
Yai T, Iwakura S, Morichi S (1997) Multinomial probit with structured covariance for route choice behavior. Transp Res Part B 31(3):195-207

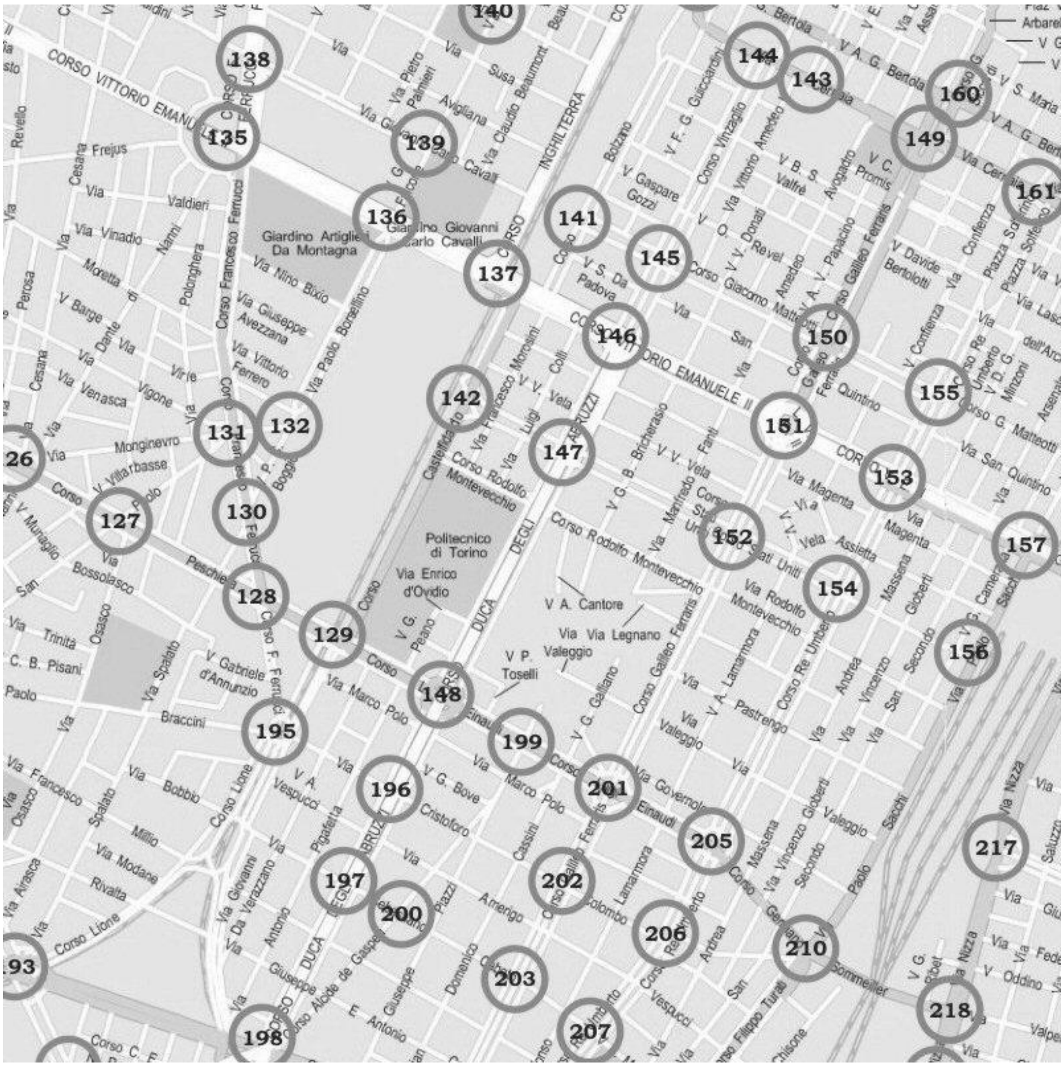

Fig. 1 An example of coded junctions on the city map

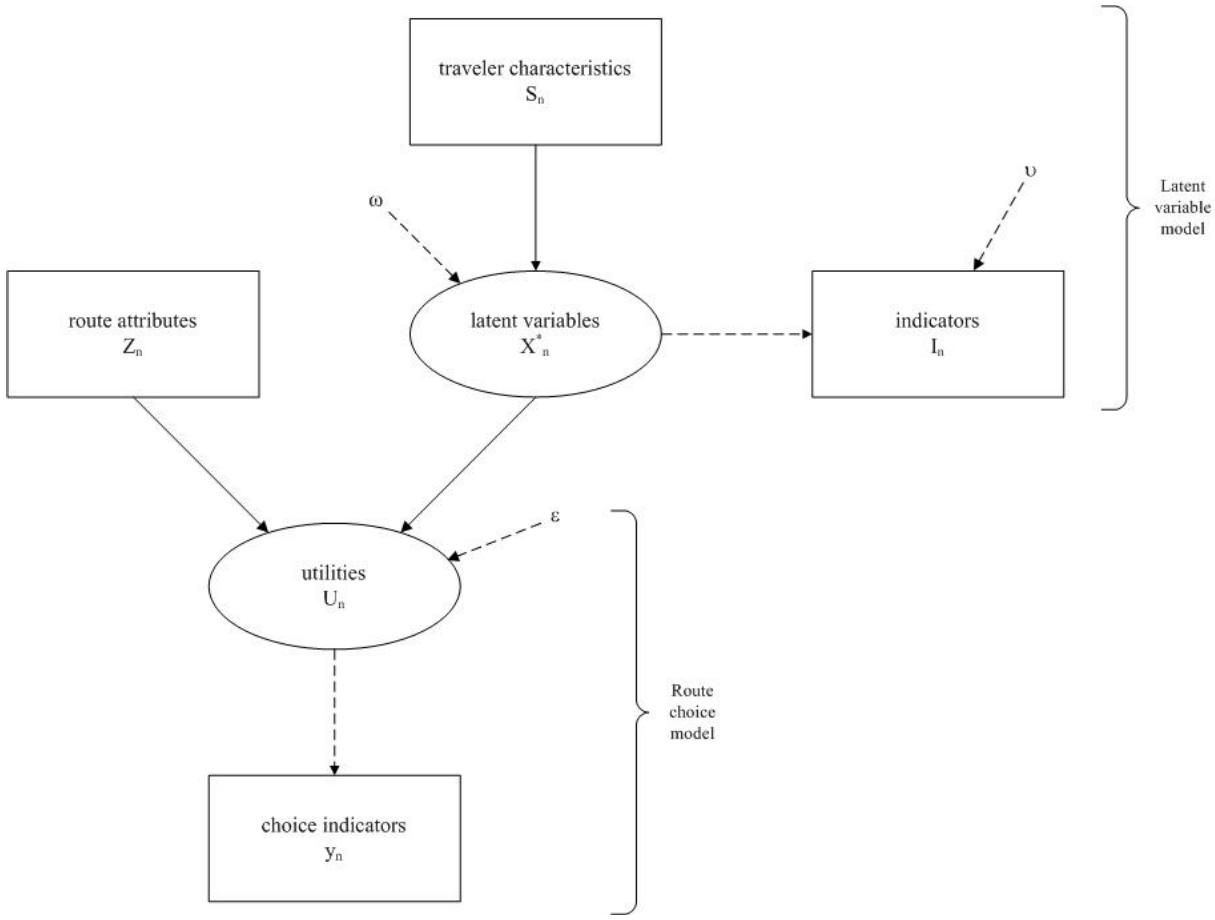

Fig. 2 Hybrid model framework

Table 1 Characteristics of the survey participants

Characteristic	Description	%
Gender	male	58.9
	female	41.1
Age	less than 25 years old	10.9
	between 25 and 35 years old	31.4
	between 35 and 45 years old	31.4
	between 45 and 55 years old 48.3	16.9
	more than 55 years old	9.4
Family composition	single	15.7
	married without children	49.2
	married with children	35.1
Education	intermediate school	3.7
	high-school	30.3
	M.Sc.	39.7
	Ph.D.	20.0
Location	residence inside the city	62.3
	residence outside the city	37.7
Stops	usually stops on the way to work	27.4
	never stops on the way to work	72.6

Table 2 Latent variables and indicators considered in the hybrid model

Latent variable	Indicator	Description
MEM mnemonic ability	I_1 - MEMROUTE	Remembering a route just learned
	I_2 - MEMHOME	Remembering the positions of objects at home
	I_3 - MEMMIND	Remembering dates and events
	I_4 - MEMLAYOUT	Remembering the layout of a shopping mall
	I_5 - MEMWAY	Remembering a route as a passenger
	I_6 - MEMPARK	Remembering the parking place
HAB habit within the choice environment	I_7 - SMRTWORK	Using the same route to go to work
	I_8 - SMRTSHOP	Using a route just learned
	I_9 - DISTSHOP	Shopping daily close to home or work
	I_{10} - SAMESHOP	Shopping daily in the same place
	I_{11} - TENDSDCH	Tendency to change route for traffic conditions
	I_{12} - TENDCHSG	Tendency to change route for a suggestion
FAM familiarity with the choice environment	I_{13} - DSCFAMRT	Describing familiar routes
	I_{14} - DSCRTHOM	Describing the route to own house
	I_{15} - EVALROUTE	Evaluating time for a generic route
	I_{16} - NAVHOME	Navigating in the dark at home
	I_{17} - DRVMAIN	Driving through main roads
SPAB spatial ability	I_{18} - BUYMAP	Buying a map in an unknown city
	I_{19} - TENDMAP	Tendency to use a map
	I_{20} - DISTTOWN	Evaluating distances on a map
	I_{21} - DRVLANDM	Driving through landmarks
	I_{22} - DRVSCEN	Driving through scenic roads
TSAV time saving skill	I_{23} - ESTTIME	Estimating time for the route to own house
	I_{24} - USEINT	Using internet for information search
	I_{25} - SHORTCUT	Looking for shortcuts on a generic route
	I_{26} - DRVNOTL	Driving on roads without traffic lights
	I_{27} - TNDESTTM	Tendency to estimate time correctly
	I_{28} - TNDESTDS	Tendency to estimate distances correctly

Table 3 Estimates of the measurement equations of the latent variable model

	MEM	
Variable	estimate	t-stat.
MEMROUTE	1.000	-
MEMHOME	0.969	6.43
MEMMIND	0.869	6.31
MEMLAYOUT	0.746	5.32
MEMWAY	1.346	7.09
MEMPARK	1.338	7.72

	HAB	
Variable	estimate	t-stat
SMRTWORK	1.000	-
SMRTSHOP	0.989	2.25
DISTSHOP	1.485	3.12
SAMESHOP	1.724	3.12
TENDSDCH	-0.958	-2.11
TENDCHSG	-1.401	-3.65

	FAM	
Variable	estimate	t-stat.
DSCFAMRT	1.000	-
DSCRTHOM	0.767	7.16
EVALROUTE	0.640	5.48
NAVHOME	0.186	2.45
DRVMAIN	-0.099	-1.70

	SPAB	
Variable	estimate	t-stat
BUYMAP	1.000	-
TENDMAP	0.519	2.46
DISTTOWN	2.772	2.75
DRVLANDM	0.682	2.38
DRVSCEN	-0.562	-2.78

	TSAV	
Variable	estimate	t-stat
ESTTIME	1.000	-
USEINT	3.346	2.47
SHORTCUT	3.842	2.43
DRVNOTL	3.669	2.41
TNDESTTM	2.057	2.24
TNDESTDS	2.288	2.12

Table 4 Estimates of the structural equations of the latent variable model

	MEM		HAB		FAM	
Variable	estimate	t-stat.	estimate	t-stat.	estimate	t-stat.
MALE	0.221	2.22	-0.080	-1.13	0.238	2.52
AGEL35	0.192	3.26	-0.216	-2.67	-	-
AGEM55	-0.125	-1.87	0.217	2.27	-	-
EDUC	0.270	2.61	0.223	3.00	-	-
SINGLE	-	-	-0.319	-3.15	-	-
CHILDREN	0.306	2.78	-	-	0.260	2.60
STOPS	-0.361	-3.18	0.260	3.35	-0.364	-3.40
RESCITY	-	-	0.246	3.37	0.240	2.23
CONSTANT	-0.204	-2.08	-0.125	-1.76	-0.271	-2.67

	SPAB		TSAV	
Variable	estimate	t-stat.	estimate	t-stat.
MALE	0.365	4.98	0.267	3.29
AGEL35	-	-	0.234	2.53
AGEM55	-	-	-0.315	-2.91
EDUC	0.607	7.99	0.272	3.28
SINGLE	-	-	-0.237	-2.08
CHILDREN	0.070	2.14	-	-
STOPS	-0.208	-2.54	0.228	2.60
RESCITY	-0.144	-1.91	0.247	2.03
CONSTANT	-0.461	-3.64	-0.463	-3.27

Table 5 Estimates of the route choice model

	PSC-LOGIT without latent variables		PSC-LOGIT with latent variables	
Variables	estimate	t-stat.	estimate	t-stat.
DIST	-0.620	-4.13	-0.838	-4.81
TIME	-0.341	-6.62	-0.284	-5.11
DELPC	-0.458	-3.13	-0.300	-2.91
TMRPC	0.525	3.68	0.520	3.47
TURNS	-0.163	-2.46	-0.190	-2.82
PSC	0.655	3.12	0.690	3.29
MEM – DIST	-	-	-0.533	-1.87
MEM – DELPC	-	-	-0.137	-1.80
HAB – DIST	-	-	0.893	2.59
HAB – TMRPC	-	-	-0.810	-2.31
HAB – TURNS	-	-	1.910	2.38
FAM – DELPC	-	-	0.313	1.76
FAM – TMRPC	-	-	-0.815	-2.36
FAM - TURNS	-	-	2.120	2.31
SPAB – TMRPC	-	-	0.515	2.03
SPAB – TURNS	-	-	-1.200	-2.57
TSAV – TIME	-	-	-0.135	-1.71
TSAV - DELPC	-	-	-0.145	-1.72
Number of parameters:	6		103	
Null log-likelihood:	-1298.38		-1298.38	
Final log-likelihood:	-1061.69		-947.92	
Rho-bar squared:	0.182		0.270	
Adjusted rho-bar squared:	0.178		0.191	

Table 6 Estimates of the route choice model from different choice set generation methods

Variables	PSC-LOGIT branch and bound estimate	t-stat.	PSC-LOGIT random walk estimate	t-stat.	PSC-LOGIT bfs-le estimate	t-stat.
DIST	-0.838	-4.81	-0.917	-3.45	-0.793	-4.04
TIME	-0.284	-5.11	-0.313	-4.42	-0.253	-4.21
DELPC	-0.300	-2.91	-0.340	-2.11	-0.327	-2.69
TMRPC	0.520	3.47	0.562	2.80	0.552	3.54
TURNS	-0.190	-2.82	-0.216	-2.21	-0.209	-2.59
PSC	0.690	3.29	0.656	2.14	0.673	3.21
MEM – DIST	-0.533	-1.87	-0.623	-1.86	-0.482	-1.68
MEM – DELPC	-0.137	-1.80	-0.144	-1.75	-0.160	-1.79
HAB – DIST	0.893	2.59	1.066	2.13	0.841	2.10
HAB – TMRPC	-0.810	-2.31	-0.897	-1.99	-0.890	-2.50
HAB – TURNS	1.910	2.38	2.002	1.93	1.959	2.38
FAM – DELPC	0.313	1.76	0.359	1.69	0.343	1.68
FAM – TMRPC	-0.815	-2.36	-0.941	-1.97	-0.855	-2.38
FAM - TURNS	2.120	2.31	2.196	2.07	2.285	2.16
SPAB – TMRPC	0.515	2.03	0.587	1.87	0.548	2.13
SPAB – TURNS	-1.200	-2.57	-1.252	-2.00	-1.279	-2.37
TSAV – TIME	-0.135	-1.71	-0.157	-1.62	-0.139	-1.67
TSAV - DELPC	-0.145	-1.72	-0.171	-1.60	-0.152	-1.64
Null log-likelihood:	-1298.38		-1298.38		-1298.38	
Final log-likelihood:	-947.92		-956.05		-953.90	
Rho-bar squared:	0.270		0.264		0.265	
Adjusted rho-bar squared:	0.191		0.184		0.186	