HYBRID CHOICE MODEL FOR PROPENSITY TO TRAVEL AND TOUR COMPLEXITY

Lissy La Paix
Transport Research Centre TRANSyT, Universidad Politécnica de Madrid, Av. Profesor Aranguren, 28040 Madrid, lissy.lapaix@upm.es

Andrés Monzón
Transport Research Centre TRANSyT, Universidad Politécnica de Madrid, Av. Profesor Aranguren, 28040 Madrid, andres.monzon@upm.es

Elisabetta Cherchi
DTU Transport, Technical University of Denmark, Department of Transport, Bygningstorvet 116 Vest, 2800 Kgs. Lyngby, elich@transport.dtu.dk

SUMMARY

During the last years cities around the world have invested important quantities of money in measures for reducing congestion and car-trips. Investments which are nothing but potential solutions for the well-known urban sprawl phenomenon, also called the “development trap” that leads to further congestion and a higher proportion of our time spent in slow moving cars. Over the path of this searching for solutions, the complex relationship between urban environment and travel behaviour has been studied in a number of cases. The main question on discussion is, how to encourage multi-stop tours?

Thus, the objective of this paper is to verify whether unobserved factors influence tour complexity. For this purpose, we use a data-base from a survey conducted in 2006-2007 in Madrid, a suitable case study for analyzing urban sprawl due to new urban developments and substantial changes in mobility patterns in the last years. A total of 943 individuals were interviewed from 3 selected neighbourhoods (CBD, urban and suburban). We study the effect of unobserved factors on trip frequency. This paper present the estimation of an hybrid model where the latent variable is called propensity to travel and the discrete choice model is composed by 5 alternatives of tour type.

The results show that characteristics of the neighbourhoods in Madrid are important to explain trip frequency. The influence of land use variables on trip generation is clear and in particular the presence of commercial retails. Through estimation of elasticities and forecasting we determine to what extent land-use policy measures modify travel demand. Comparing aggregate elasticities with percentage variations, it can be seen that percentage variations could lead to inconsistent results. The result shows that hybrid models better explain travel behavior than traditional discrete choice models.
1 INTRODUCTION

The phenomenon called *Urban Sprawl* is produced by the movement of population from the city centre to low density urban areas, with poorer accessibility and facilities, and as a consequence high car-dependency. City structures are changing from mono-centric to polycentric cities (Gordon, 1986; Small, 1992; Clark, 1994; McDonald, 1994; R. Cervero, 1997). This controversial term has received a lot of attention in recent years due to its association with the environment, health, transport and public investments, and to improve our understanding of the relationship between travel behaviour and urban structure (Giuliano, 1993; Handy, 1996). This phenomenon means low density developments which are more difficult and expensive to serve by more efficient transport modes. Urban Sprawl is also called the “development trap” that leads to further congestion and a higher proportion of our time spent in slow moving cars (Ortuzar & Willumsen, 2011).

Several authors have used discrete choice models to study the relation between, urban sprawl, travel demand and built environment. Recent research focus on: vehicles miles driven or VMD (Handy et al., 2005), tour-frequency (Limanond & Niemeier, 2004), shopping tour (Agyemang-Duah et al., 1995), type of activity (Naess, 2006), modal choice or modal changes (Bento et al., 2005), and trip frequency (La Paix et al., 2010; La Paix, 2010; La Paix et al., 2012). And, a notable methodological evolution is observed in the literature, from the traditional regression models to more sophisticated discrete choice models.

According to this, the existing background demonstrates the potential of latent variables to capture significant effects that cannot be captured by observable variables in the choice models. Thus, the explanation of the decision making process is improved by adding latent variables. There are several approaches with which researchers have aimed to capture intangible factors in the modelling process. However there is still a gap in the empirical application of hybrid choice models, and this gap is even larger between the current application of hybrid choice models and the relationship built environment – travel behaviour. Some authors have considered different travel dimensions, but they did not analyzed tour. The modelling of tour decisions provides an incremental improvement over trip-based model systems, incorporating an explicit representation of temporal-spatial constraints among activity stops within a tour (Bowman & Ben-Akiva, 2001). No author, so far, seems to have analyzed the relationship between neighbourhood type and tour structure with hybrid choice models.

The aim of this paper is to study the effect of built environment and socio-economic characteristics in the discrete choice among different tour structures. At the same time, following the existing literature on the effect of latent factors in the discrete choice (Ben-Akiva et al., 1999; Ben-Akiva et al., 2002; Walker and Ben-Akiva, 2002; Atasoy et al., 2010; Hurtubia et al., 2010; Yáñez et al., 2010)), it can be thought that the discrete
choice among type of tours can also be affected by unobservable attitude towards travel that is not reflected in the explanatory variables. To study this effect hybrid choice theory and model (HCM) is used, where the latent variable measures the propensity to travel of each individual, while the discrete choice is among type of tours. Both the latent variable and the discrete choice are functions of the individual socio-economic and land use characteristics, but the tour choice is also a function of travel characteristics and a function of the latent variable. In this way it was possible to measure the effect of the land use and socio-economic characteristics have on the choice of tours directly and indirectly through the propensity to travel.

Due to the increasing importance of latent variables on the transport research, it can be thought that some unobserved factors could influence the choice among type of tour. This paper is an analysis of, first, the propensity to travel and, after, the discrete choice.

1.1 Definitions of Propensity and Tour

Firstly, it is important to define propensity and tour. Propensity refers to the natural or acquired tendency, inclination, or habit in a person or thing. This might be thought of as a general willingness to do something, which, at the same time, influences how frequently an action is carried. In the psychological literature “Propensity” is often used with specific meanings, such as “Propensity to Trust” others (Mayer, 1995), “Risk Propensity” as the tendency of a decision maker either to take or to avoid risks” (Sitkin, 1992) or “Propensity effect” as a reversal of the traditional hindsight bias (Roese, 2006). Following this definition, in the present paper, the propensity to travel (PT) is defined as the individual tendency to travel, and it is measured by the daily trip frequency.

In this context, “Tour” is defined as a sequence of trip segments during a full-day. In particular the tour structures were defined in terms of type of main activities of the tour and number of stops realised during the tour for other purposes other than the main activity.

Fig. 1 shows the variables interacting in the representation of this phenomenon. Variables in grey indicate the information only available once the tour starts; these variables are related to tours and they are also called “Level of Service” (LOS). Variables in white background represent the information that is available for forecasting purposes, and can be used as predictors for the propensity to travel.
The main motivation to select the propensity to travel as latent variable is because it is possibly that an unobservable attitude could affect discrete choice model, but it is not reflected in the explanatory variables. As for the discrete choice, the type of tour was selected because it incorporates an explicit representation of temporal-spatial constraints among activity stops within a tour.

2 METHODOLOGY

The present work follows the general framework and methodology reported in Walker (2001) for incorporating latent variables into choice models via the integration of the choice and latent variable models. The integrated choice and latent variable structure explicitly models the latent variables that influence the choice process. Fig. 2 shows the framework of the integrated latent variable and choice model used to study the relation between BE and tour choice. It was assumed that socioeconomic (SE) and built environment characteristics (BE) play a role in both the latent and discrete choice model, while the travel attributes (LOS) only affect the choice of the type of tour, because this information only exists once a tour has started. Before arriving to the final specification of the HCM different specifications were tested. The observable indicator is the number of trips that links propensity to travel with the tour complexity.
3 ANALYTICAL FRAMEWORK

As in the general framework previously discussed, we need the distribution of the latent variable, the number of trips, given the observed SE and BE attributes. Let PT be the latent variable, and \( S_n \) and \( Z_n \) two vectors of explanatory variables respectively for the SE and LU characteristics. The structural equation for PT is specified as follows:

\[
PT_n = \sum_k \lambda_k^S S_{kn} + \sum_m \lambda_m^Z Z_{mn} + \omega_n
\]  

Where \( PT_n \) is the propensity to travel for individual \( n \). \( S_n \) is a vector of SE characteristics with \( k \) elements, \( Z_n \) is the vector of BE attributes with \( m \) elements, \( \lambda^S \) and \( \lambda^Z \) are two vectors of parameters associated respectively to the SE and LU characteristics, while \( \omega_n \) is the error term Normal distributed with zero mean and standard deviation \( \sigma_\omega \).

We also need the distribution of the indicators conditional on the values of the latent variable. The measurement equation for propensity to travel (\( PT_n \)) is specified as follows:

\[
I_{ln} = \alpha_l PT_n + \nu_{ln}
\]

where \( I_{ln} \) is l-th indicator of the latent variable (\( PT_n \)), \( \alpha_l \) is the associated parameter to be estimated and \( \nu_{ln} \) is the error term, Normal distributed with zero mean and standard deviation \( \sigma_\nu \).

Similarly to the latent variable model, also for the discrete choice model, we need the distribution of the utilities (\( U_{jn} \)) that individual \( n \) associates with each type of tour \( j \) and a measurement equation to identify the choice. The utility function is expressed as a function
of a vector of socioeconomic attributes \((S_n)\), a vector of BE characteristics \((Z_n)\), a vector of attributes of the tour \((LOS_{nj})\) and the propensity to travel \((PT_n)\) of each individual \(n\).

\[
U_{jn} = ASC_j + \sum_i \beta_i^z S_{ln} + \sum_r \beta_r^Z Z_{nr} + \sum_s \beta_s^{LOS} LOS_{njs} + \beta_j^{PT} PT_n + \varepsilon_{jn} \tag{3}
\]

Note that the discrete choice and the latent variable models can include different attributes, hence the vectors \(S_n\) and \(Z_n\) can be different between equations (6-1) and (6-3). The parameters associated with these attributes in the discrete choice that are, of course, different from those in the latent variable. \(\varepsilon_{jn}\) is the error term extreme value distributed with mean zero and \(\sigma_e\) standard deviation.

The choice set in the alternative specifications involves 5 alternatives, and are represented by the following types of tours:

- \(j=1\) is HOME: no trips during the day.
- \(j=2\) is a Home-Work-Home tour (HW/SH), which includes
  - Simple tour from home to work and back
  - Simple school tour from home to school and back
- \(j=3\) is a Work/school tour with at least 1 additional stop for another activity (HW/SH+), and it includes:
  - Work tour with an intermediate stop at home.
  - Work tour with an intermediate stop at home, plus 1 or more additional stops.
  - Work tour with a work-based sub-tour, and any number of additional stops
- \(j=4\) is a Simple tour from home to shopping and back (HOH/SHOP) or a Simple tour with purpose other than work or school or shopping
- \(j=5\) is a Shopping tour with at least 1 additional stop for another activity (HOH+/SHOP+) or a tour with purpose other than work or school, with at least 1 additional stop for another activity

The choice model was built assuming extreme value distribution for the error terms of the alternatives; hence the probability of individual \(n\) choosing the alternative \(j\) is the probability of choosing the alternative conditional on the observed and unobserved variables:

\[
P(j_n|S_n, Z_n, LOS_{nj}, PT_n, \beta^{S,Z,LOS,PT}, \sigma_e) = \text{Prob} \left[ U_{jn} \geq U_{ln}, \forall \in C_n \right] \tag{4}
\]

Where \(C_n\) is the choice set of the individual \(n\). Since PT is unknown, then the probability of individual \(n\) choosing alternative \(j\) is the integral:
4 ESTIMATION

Two approaches can be used to estimate the hybrid model: sequential and simultaneous. Simultaneous estimation is applied for estimating in this paper, because it leads to more efficient estimates (M. Ben-Akiva et al., 1999). In hybrid model the latent variable and the discrete choice are estimated jointly. In the integrated model, we have to estimate the joint probability of observing both the choice \( j \) for individual \( n \) and the latent variable \( PT_n \). Therefore, the joint probability of choosing a tour \( j \) and to observe the indicator \( I \), is given by:

\[
P(j, I_n | S_n, Z_n, LOS_{nj}, PT_n; \beta^{SZ,LOS,PT}, \alpha_t, \lambda^{SZ}, \sigma_e, \sigma_v, \sigma_\omega)
\]

\[
= \int_{PT_n} P(j | S_n, Z_n, LOS_{nj}, PT_n; \beta^{SZ,LOS,PT}, \sigma_e) f_I(I_n | PT_n; \alpha_t, \sigma_v) f_{PT}(PT_n | S_n, Z_n; \lambda^{SZ}, \sigma_\omega) dPT
\]

Where the densities of \( PT_n \) and \( I_n \) are given respectively by:

\[
f_{PT_n}(PT_n | \sigma_\omega) = \frac{1}{\sigma_\omega} \phi \left( \frac{PT_n - \sum_k \lambda_k^{SZ} z_k - \sum_m \lambda_m^{PZ} z_m}{\sigma_\omega} \right)
\]

\[
f_I(I_n | \sigma_v) = \frac{1}{\sigma_v} \phi \left( \frac{I_n - \alpha_{PT_n}}{\sigma_v} \right)
\]

The maximum likelihood is obtained, as always, from maximizing the logarithm of the likelihood function \((L)\) over the unknown parameters:

\[
L = \sum_n \sum_{i \in C_n} d_{jn} \log P(j, I_n | S_n, Z_n, LOS_{nj}; \beta^{SZ,LOS}, \alpha^{PT}, \sigma_e, \sigma_v, \sigma_\omega)
\]

Where the binary variable \( d_{jn} \) characterizes the individual decisions and it is defined as:

\[
d_{jn} = \begin{cases} 1 & \text{if } U_{jn} > U_{in}, \forall \in C_n \\ 0 & \text{in other case} \end{cases}
\]

and \( C_n \) is the choice set of each individual, i.e. the set of alternatives available for each individual.
5 RESULTS

The main findings that emerge from this paper are: firstly, unobserved factors play an important role in travel behaviour, and more specifically, on the discrete choice of type of tour. And secondly, BE is a key observed issue in those intangible factors, because neighbourhood characteristics had significant influence in the propensity to travel. The main motivation to select the propensity to travel as a latent variable is because it was thought that unobservable attitudes could affect the discrete choice model, but it is not reflected in the explanatory variables. The PT for a single individual measures in fact how frequently s/he travels depending on her/his socioeconomic attributes and the characteristics of the neighbourhoods where s/he lives and performs other activities during the tour. Since the effect of PT is significant on tour complexity, there is partially an explanation from individual’s preferences. This part of the explanation is not observed but it is manifested and, as consequence, it is controllable.

Although the hybrid model does not directly include mixed land-use effect, the effect of the neighbourhood type suggest that mixed land-uses do not encourage a propensity to travel and, as a consequence trip generation, while positively influences tour complexity and thus stop making propensity.

The results of this paper also reveal that higher densities favour tour complexity, and significantly influence propensity to travel, and in turn this propensity positively influence tour complexity. This adds new evidence to the current literature, where often opposite effects of the mixed land-use and density on travel behaviours are reported. For example, Limanond and Niemeier (2004) reveal that land use patterns have no impact on the whole shopping tour frequency; Kitamura et al. (1997) suggest that land use policies promoting higher densities and mixtures may not alter travel demand materially unless the attitudes of residents are also changed. By contrast, Cervero and Kockelman (1997) find that land-use diversity reduce trip rates and encourage non-auto travel; Cervero (1996) reports that mixed-use development is more important than density in affecting non-motorized work trip mode shares while Kockelman (1997) reports that density has a negligible impact on travel behaviour (except with respect to auto ownership) once accessibility is taken into account. A detailed review of these effects is reported in Badoe and Miller (2000).

5.1 Marginal utilities

If we look at the level of service measure, the travel time has been included only in the utility function of non-working tours-with-stops, because people travelling for non-working purposes should have higher marginal utility of travel time than people travelling for other purposes. We found that, the parameter is negative, which confirm that people who travel longer distances during the day are less likely to carry out other intermediate activities or tours with stops. It shows also the link between time constraints and an individual’s decision process, because longer travel time entails less time for intermediate
Variable & Type of tours & \\
& HWH & HWHs & HOH & HOHs \\
Age 14-21 & 1.69 & 4.00 & 2.75 & 4.09 \\
Age 22-39 & 0.40 & 2.72 & 1.06 & 1.88 \\
Age 40-49 & -0.78 & 1.56 & 1.07 & 2.44 \\
Age 50-64 & -0.80 & 2.34 & 2.22 & 3.58 \\
Female married & & 2.40 & 1.10 & 2.50 \\
Worker & & 2.00 & 4.34 & 1.07 & 2.44 \\
Own car & & 0.02 & 0.01 & 0.02 \\
Single family & & -2.32 & -1.06 & -2.42 \\
Condominium & & -2.29 & -1.05 & -2.39 \\
Metro600 & & -0.11 & -0.05 & -0.13 \\

Table 1 Marginal utilities accounting for the effect of the propensity to travel

5.2 Aggregate elasticities

Finally, Table 2 includes the aggregate elasticity for each type of tour, computed for both the hybrid and the simple discrete choice model. As expected, the ratio of workers between origin and destination has the highest impact in the elasticity and it has the same sign in both HCM and MNL models. This is because the attribute “RatioWorkers” is not included in the latent model, so its effect is similar for both the hybrid and the simple multinomial logit model. Note also that, although the sign of the “RatioWorkers” coefficients is negative in all the three alternatives (HWH, HWHs, HOH), the elasticity for HWHs is positive. This is because the ratio of workers is a socioeconomic attribute and it does not vary among tours, so a variation of the ratio of works directly affects all the three alternatives.

Table 2 Average Elasticity for Ratio of Workers and Travel Time

The number of metro stations instead is included in both the latent variable (alone) and in the discrete choice among tour (summed to the number of bus stops), so the elasticity is of course different in the two model formulations. In the MNL, the attribute is included only
in the HOHs tour, so the elasticity is negative for the HOHs tours (because the attribute has a negative coefficient) and it is positive and equal in all the other alternatives. In the HCM instead, the elasticity is affected by the effect of the number of stations in the latent propensity to travel, and it is in fact negative for the three alternatives where the latent variable is included. Moreover, because of the strong effect of the latent propensity to travel, the effect of the number of metro stations in the HCM is also higher (in absolute value) in the HMC than in the MNL model. This reinforces the importance to properly account for the latent propensity to travel into the discrete choice among tours.

5.3 Percentage Variations

To evaluate the forecasting capability of the estimated hybrid choice model, we computed the variation in the aggregate market shares for a couple of simple policy measures. The response to a change in the prediction was calculated as the percent change in the aggregate share of mode $j$ over the initial situation (do-nothing):

$$\Delta P_j = \frac{P_j - P_j^0}{P_j^0}$$

where $P_j^0, P_j$ are the aggregate probabilities of choosing mode $j$ before (do-nothing) and after introducing the measure, calculated by sample enumeration.

Table 3 shows the variation in market shares for each type of tour, after assuming a 40% increase, uniform across all individuals, for the ratio of workers, for the number of public transport stops and travel time. Each attribute was increased once at a time and the market share variation computed accordingly. Table 3 reports the market share variation computed with both the hybrid and the simple discrete choice model, in order to evaluate the effect in prediction of accounting for latent effects.

An interesting discussion could be approached about results from Table 3. Comparing aggregate elasticities with percentage variations, it can be seen that percentage variations could lead to inconsistent results. Similarly, DCM tends to overestimate the elasticity, and this is consistent with previous findings. However, exceptions take place depending on the estimator’s efficiency in the market share.

As expected, the ratio of workers between origin and destination has the highest impact in the variation of the market share. In line with the model results, people perform fewer tours (except tours for other purposes and many stops) and stay more at home. As can be seen in Table 3, comparing simple discrete and hybrid choice model, the results are mixed. On one hand, simple choice model lead to higher variations in market shares for HOHs tours. By contrast, lower variations are obtained for travel time in almost all types of tour. And, lower variations are obtained from ratio of workers in home and HWH.
### Table 3 Percentage Variations and Market Share

<table>
<thead>
<tr>
<th>Type of tours</th>
<th>40% increases in the Ratio of Workers (O/D)</th>
<th>40% increases in # of Metro Stations</th>
<th>40% increases in Travel time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HCM</td>
<td>DCM only</td>
<td>HCM</td>
</tr>
<tr>
<td>Home</td>
<td>220.32%</td>
<td>129.80%</td>
<td>1.44%</td>
</tr>
<tr>
<td>HWH</td>
<td>-67.33%</td>
<td>-62.51%</td>
<td>1.60%</td>
</tr>
<tr>
<td>HWHs</td>
<td>-3.26%</td>
<td>94.15%</td>
<td>-2.41%</td>
</tr>
<tr>
<td>HOH</td>
<td>-77.29%</td>
<td>-80.00%</td>
<td>-0.83%</td>
</tr>
<tr>
<td>HOHs</td>
<td>63.24%</td>
<td>127.03%</td>
<td>-2.14%</td>
</tr>
</tbody>
</table>

### 6 CONCLUSIONS AND POLICY RECOMMENDATIONS

In global terms, the model developed here indicates that perceptions and attitudes are as important as built environment. There are unobserved constructs in the analysis of travel behaviour, and the results indicate that urban planners must revise the application of urban demand models. Thus, demand models must consider latent effects.

Variations in market shares are highly associated with the latent constructs. And, policy makers must work on the acceptability of measures, in order to reach first optimum, i.e. public participation and diffusion. In this sense, findings from this paper confirm that latent variables play an important role in the definition of implementation path of transport measures.

The analysis of elasticity, presented in this paper, show that cross elasticities of travel-time is lower than the elasticity of ratioworkerOD, indicating that people are more elastic for the changes in BE than in travel time. Thus, in order to promote multi-stops tours, it is important to increase commercial activity at destination places (i.e. work-places, and leisure). The results show that people undertake more stops at destination place than at origin. The impact of the employed population in working tours is higher than travel time with stops.

Today, urban planners must promote multi-stop tours, because it represents savings in energy consumption, vehicle miles driven and time. Since travel is derived from the necessity to participate in activities, users tend to reduce travel time as much as possible. Thus, policy makers can reduce travel time by improving level of service (i.e. frequency, cost and accessibility) or by increasing proximity to destinations (i.e. number of facilities per square kilometres).
REFERENCES


La Paix, L. (2010). Relación entre la generación de viajes y la densidad urbana: Caracterización de la movilidad con modelos probit ordenados. IX Congreso Ingeniería del Transporte CIT. 5 -7 July, 2010. Madrid, Spain


ORTUZAR, J. D., & WILLUMSEN, L. G. (2011). In John Wiley & Sons (Ed.), Modeling transport. Chichester, UK:


