ASSESSMENT METHODOLOGY APPLIED TO DEMAND MEASURES TO CHANGE URBAN MOBILITY BEHAVIOR

DOCTORAL THESIS

Julio Comendador Arquero
Ingeniero de Caminos, Canales y Puertos
Máster en Sistemas de Ingeniería Civil

Madrid, 2015
UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE CAMINOS, CANALES Y PUERTOS

DOCTORAL THESIS

ASSESSMENT METHODOLOGY APPLIED TO DEMAND MEASURES TO CHANGE URBAN MOBILITY BEHAVIOR

Julio Comendador Arquero
Ingeniero de Caminos, Canales y Puertos
Máster en Sistemas de Ingeniería Civil

SUPERVISORS:
MARÍA EUGENIA LÓPEZ-LAMBAS
Doctora en Derecho
ANDRÉS MONZÓN DE CÁCERES
Dr. Ingeniero de Caminos, Canales y Puertos
Abstract

In order to minimize car-based trips, transport planners have been particularly interested in understanding the factors that explain modal choices. Transport modelling literature has been increasingly aware that socioeconomic attributes and quantitative variables are not sufficient to characterize travelers and forecast their travel behavior. Recent studies have also recognized that users’ social interactions and land use patterns influence travel behavior, especially when changes to transport systems are introduced; but links between international and Spanish perspectives are rarely dealt with.

The overall objective of the thesis is to develop a stepped methodology that integrate diverse perspectives to evaluate the willingness to change patterns of urban mobility in Madrid, based on four steps: (1st) analysis of causal relationships between both objective and subjective personal variables, and travel behavior to capture pro-car and pro-public transport intentions; (2nd) exploring the potential influence of individual trip characteristics and social influence variables on transport mode choice; (3rd) identifying built environment dimensions on travel behavior; and (4th) exploring the potential influence on transport mode choice of extrinsic characteristics of individual trip using panel data, land use variables using spatial characteristics and social influence variables.

The data used for this thesis have been collected from a two panel smartphone-based survey (n=255 and 190 respondents, respectively) carried out in Madrid. Although the steps above are mainly methodological, the application to the area of Madrid allows deriving important results that can be directly used to forecast travel demand and to evaluate the benefits of specific policies that might be implemented in the area. The results demonstrated, respectively: (1st) transport policy actions are more likely to be effective when pro-car intention has been disrupted first; (2nd) the consideration of “helped” and “voluntary” users as tested here could have a positive and negative impact, respectively, on the use of public transport; (3rd) the importance of density, design, diversity and accessibility underlying dimensions responsible for land use variables; and (4th) there are clearly different types of combinations of social interactions, land use and time frame on travel behavior studies.

Finally, with the objective to study the impact of demand measures to change urban mobility behavior, those previous results have been considered in a unique way, a hybrid discrete choice model has been used on a 5th step. Then it can be concluded that urban mobility behavior is not only ruled by the maximum utility criterion, but also by a strong psychological-environment concept, developed without the mediation of cognitive processes during choice, i.e., many people using public transport on their way to work do not do it for utilitarian reasons, but because no other choice is available. Regarding built environment dimensions, the more diversity place of residence, the more difficult the use of public transport or walking.

Key words: pro-car intention, pro-public transport intention, social capital influence, urban built environment variables, smartphone-based panel survey, discrete choice models, structural equation models, transport policy measures.
Graphical Abstract

CASE STUDY: Two panel smartphone-based survey of Madrid

1st step

Structural Equation Models (SEM)

2nd step

Discrete Choice Models (DCM)

3rd step

Built Environment factors

4th step

5th step

SEM + DCM = HYBRID MODEL

"Conclusions about mobility behavior"
Acknowledgements (in Spanish)

Durante estos años son muchas las personas que han participado en este trabajo y a quienes quiero expresar mi gratitud por el apoyo y la confianza que me han prestado de forma desinteresada.

Gracias a mis directores de tesis. Desde que entré con una beca de verano en TRANSyT hace más de 7 años nunca me han faltado oportunidades en el mundo de la investigación, a pesar de coincidir con los años de crisis económica. Imagino que es porque siempre habéis confiado en mí. Esto sólo es un punto y seguido porque seguiremos colaborando.

Gracias a Floridea Di Ciommo. Si hoy soy alguien en el buscador “Google Scholar” es gracias a ti. Me he permitido el lujo de conocerte y me tranquiliza saber que existe el buen corazón en las personas.

Rico Wittwer. Gracias por recordarme que la amistad entre hombres también es posible. No sabía yo que avanzar en esta tesis podía ser tan divertido.

Gracias al resto de profesores que han sido coautores de alguno de los artículos en los que se fundamenta esta tesis: Juan Carlos García-Palomares, Elisabetta Cherchi y Juan De Dios Ortúzar. Uno siempre aprende de sus mayores.

Tampoco me puedo olvidar de mis jefes en la estancia predoctoral que hice en la Universidad de California (Santa Bárbara): Kostas Goulias y Srinath Ravulaparthy. En lo práctico me permitisteis usar vuestras investigaciones para aplicarlas en esta tesis; y en lo humano me abristeis las puertas a un nuevo mundo.

A Carmen y al resto de compañeros de TRANSyT. Coincidi con tantos que resumo agradeciéndoos todos los ratos de café y sobre todo de cervezas. Eso de que los investigadores son “ratas de biblioteca” no va con vosotros.

Gracias a mis hermanos y al resto de amigos de Toledo. Os ha tocado sufrir mis penas por esta tesis. Al menos os corresponden un par de páginas a cada uno.

A la mujer que me trajo a este mundo. Gracias por esas comedas envasadas, por hacerme los recados, por madrug conmigo, por quitar hierro al asunto ante las dificultades,… por hacer de madre. Papá se lo pasa en grande desde el cielo cuando te ve de un lado para otro.

Y gracias a ti. La Providencia ha querido que seamos doctores a la vez. Seguiremos cerrando etapas de la mano.
Publications from this thesis

This thesis contains five studies:


The publishers of the three last papers hold the copyright for that content, and access to the material should be sought from the respective journals.
## Index

### CHAPTER 1: INTRODUCTION. OBJECTIVES

Subchapter 1.1 – Urban mobility: theoretical background of methods ........................................... 3  
Social and land use variables ........................................................................................................ 4  
Methodologies .................................................................................................................................. 5  
Panel surveys and GPS technology ................................................................................................. 6  
HABIT project .................................................................................................................................. 8  
Hypotheses ....................................................................................................................................... 9  
Definitions ........................................................................................................................................ 10  
Subchapter 1.2 – Objectives and structure of the thesis ................................................................. 10  
Subchapter 1.3 – Analysis of social and land use effects as pro-change factors in travel behavior .. 12  
Social capital influence on travel behavior ....................................................................................... 12  
The spatial dimension on travel behavior ......................................................................................... 13  
Subchapter 1.4 – Assessment methodology for this thesis .............................................................. 14  
Structural equation models ............................................................................................................. 14  
Discrete choice models ................................................................................................................... 16  
Hybrid discrete choice models ........................................................................................................ 17  

### CHAPTER 2: SMARTPHONE-BASED PANEL SURVEY

Panel survey design aspects ............................................................................................................. 23  
Characteristics of Madrid’s transport system .................................................................................. 24  
Subchapter 2.1 – Pilot survey: a Revealed Preferences survey of New Stations Users (RP-NSU)... 25  
The RPNSU: Design and Descriptive analysis ................................................................................. 26  
Econometric and statistical assumptions for the treatment of a RP-NSU ..................................... 28  
Model definition and results ........................................................................................................... 30  
The importance of subjectivity on travel behavior ........................................................................ 33  
Subchapter 2.2 – Smartphone-based panel survey design. Testing smartphone technology ......... 34  
General aspects ............................................................................................................................... 34  
Sample size determination ............................................................................................................... 35  
The smartphone-based panel survey and its evolution ................................................................. 35  
Subchapter 2.3 – Data check and descriptive analysis .................................................................... 38
CHAPTER 3: STRUCTURAL EQUATION MODELS APPROACH .......................................................... 41
  Subchapter 3.1 – Understanding travel behavior to change intentions towards private car and public transport ......................................................................................................................... 43
  Exogenous and endogenous variables on travel behavior studies ............................................. 44
  Formulation of SEM and mediation effect ............................................................................. 47
  Definition of variables and results .......................................................................................... 49
  Brief of principal travel behavior relationships ...................................................................... 54
  Subchapter 3.2 – Urban built environment analysis ................................................................. 55
  The relevance of built environment ...................................................................................... 55
  How to measure spatial dimension? ....................................................................................... 55
  Formulation of Factor Analysis ........................................................................................... 57
  Definition of variables and results ........................................................................................ 58
  Final urban built environment factors .................................................................................. 61

CHAPTER 4: DISCRETE CHOICE MODELS APPROACH ............................................................ 63
  Subchapter 4.1 – Exploring the role of social capital influence variables on travel behavior ....... 65
  The relevance of social capital influence .............................................................................. 65
  Definition of variables and results ........................................................................................ 66
  New ways to study social capital influence on travel behavior ............................................. 68
  Subchapter 4.2 – Built environment and Social Capital variables influences on travel behavior .... 69
  Formulation of DCM and panel effect .................................................................................. 70
  Definition of variables and results ........................................................................................ 71
  Time evolution, social capital and space for exploring travel behavior ................................... 75

CHAPTER 5: HYBRID MODEL APPROACH ................................................................................. 77
  Introduction ............................................................................................................................ 79
  Formulation of sequential hybrid discrete choice model ..................................................... 80
  Definition of variables and results ....................................................................................... 80
  Testing the effect of psychological and built environmental latent variables ......................... 84
CHAPTER 6: CONCLUSIONS AND RECOMMENDATIONS ................................................................. 87
    Subchapter 6.1 – Discussion of results ............................................................................. 89
        Theoretical contributions .......................................................................................... 89
        Methodological contributions .................................................................................. 90
    Subchapter 6.2 – Policy implications .......................................................................... 91
    Subchapter 6.3 – Future researches ............................................................................. 92
REFERENCES ...................................................................................................................... 95

ANNEX ............................................................................................................................. 111

GLOSSARY .......................................................................................................................... 123
CHAPTER I:

INTRODUCTION. OBJECTIVES
Subchapter 1.1 – Urban mobility: theoretical background and methods

A known pro-car mobility paradigm by Scheller and Urry (2006) defines the automobile as being “experienced through a combination of senses and sensed through multiple registers of motion and emotion”. There is a positive pro-car intention that causes negative externalities. For example, in Spain more than half of all work trips are car-based, and car trips are increasing at a faster rate than journeys made by public transport (Monzón et al., 2011 and 2013).

Reducing the concomitant effects (i.e. pollution, congestion, accidents, stress and so on) generated by extremely high levels of mobility and by an indiscriminate use of private cars is one of the key problems debated worldwide. Apart from these problems, cities have tried to improve the operation of their transport systems by implementing various types of policies, the success of which depend on their long-term transport planning abilities and the capacity to forecast users’ reactions when confronted with changes to the system. To avoid the costly effects of taking wrong decisions it is vital to identify users’ necessities and their possible responses under the implementation of new policies. And, on the way round, how the system should be modified to induce travellers to change modes. Thus, demand estimation is a key component of any appropriate urban transport planning framework.

Most travel demand models to date have been based on cross-sectional data. This data structure allows one to get information about users’ choices in a particular moment, but fail to correctly ascertain how choices vary over time (both short and long-term variability). It is well known that the single trip observed with given characteristics at a specific moment of the day is usually only one piece of a broader choice that involves the daily, and often weekly, pattern of activities. Travel choices, and in particular modal choice, are elements in the chain of the daily/weekly activity pattern. Individual behaviour and choices might then vary over days of the same week simply because people’s needs and desires vary based on the daily/weekly activity program planned. This variability is often called day-to-day variability, or short-term or micro-dynamic effects (Clarke et al., 1982). At the same time, individual behaviour might also change due to adjustments to external changes (both changes in transport supply and in the socio-economic characteristics of the population) and are often associated to long-term variability or macro-dynamic effects, because the adjustment cannot be instantaneous and usually takes time.

Estimating how travel demand varies due to changes in supply is one of the main reasons why demand models are used in practice. However, models estimated on cross-sectional data do not allow us to study how the system evolves. Dynamic discrete choice models allow one to study both the spurious and the true structural dependence among choices made in different periods (Williams and Ortúzar, 1982). Spurious dependence refers to the existence of non-observed heterogeneity so that, once it is conditioned on it, choices in different periods of time are independent. This is usually modelled allowing for fixed or random effects (Train, 2009). On the other hand, the true structural dependence is related with the choice in different periods throughout a stochastic process (i.e. renovation models, Markov models). All these phenomena are recognized in the economic literature and mainly addressed in marketing (Keane, 1997) but they have not been appropriately treated in transport economics.

Summarizing, the primary concern for transport planners is therefore to understand: (1) travel behavior in order to minimize the attitudes that lead to pro-car intention; and (2) the adaptation strategies adopted in response to the introduction of travel demand management measures such as prohibiting car traffic from entering the city center, individualized public-transport marketing, road pricing, intelligent transport systems, carpooling, etc. To minimize these car-based trips, transport planners have been particularly interested in understanding the factors that explain modal choices. Changes in mode choice can be
explained by a mixture of quantitative factors (travel cost and time, income variation and demographical variables, such as age, social status and household size); contextual attributes (i.e. trip purpose) may also be important. In the transport model literature there has been an increasing awareness that socio-economic attributes and quantitative variables (e.g. trip cost and time) are not sufficient to characterize travelers and forecast their travel behavior. As a consequence, the literature is becoming more and more interested in investigating new variables and has incorporated attitudes and awareness in transport models (Ben-Akiva et al., 2002; Cantillo et al., 2007; Cherchi and Manca, 2011; Galdames et al., 2011; Prashker, 1979; Rieser-Schüssler and Axhausen, 2012; Yáñez et al., 2010).

Nowadays, it can be talking about traveler attitudes (perception of attributes and availability of private transport) and travel behavior at the same level because it has been shown to correlate in numerous studies since the 1970s (Recker and Golob, 1976; Dobson and Tischer, 1976 and 1978); and more recently (Verplanken et al., 1994; Domarchi et al, 2008; Eriksson et al., 2008), providing these last a better knowledge about the nature of the interrelationships between traveler attitudes and travel behavior. This aspect is extremely important because if attitudes cause behavior, then the mode choice and other traveler decisions can be influenced by changing traveler viewpoints toward public transit, carpools, and single-occupant automobiles.

**Social and land use variables**

Recent studies have also recognized that users’ social interactions –as well as their perceptions– may influence travel behavior, especially when changes to transport systems are introduced (Ben-Akiva et al., 2012; Brock and Durlauf, 2003; Carrasco and Miller, 2006). Past transport research has led to an increased understanding of behavioral processes; but, key social influence variables are rarely included directly in travel behavior models because of the difficulty in measuring the degree of integration of people with respect to their spatial proximity and social environment (social capital influence). However, the social environment (i.e. family, friends, and market) may influence travel behavior to a certain extent, and integrating it in transport behavior models can help to explain why two apparently identical individuals can make completely different decisions when facing an objectively equal situation.

Due to the increasing awareness of the influence of social variables on travel behavior, a wide range of analyses evidence that individual outcomes are linked to social interaction factors, especially within social groups defined by geographic proximity (Brooks-Gunn, 1993; Glaeser et al., 1992; Bauer and Zimmermann, 1999). Ostrom (1998) modified the rational utility theory to include social interaction variables like reciprocity, trust and reputation, obtaining better results. Methodologically, Brock and Durlauf (2001) and Skrondal and Rabe-Hesketh (2003) have already expanded the effort to employ social interaction variables in a structural estimation with binary choice models. But theoretically, Di Ciommo (2003) and Granovetter (2005) introduced the novelty term social capital to refer the social embedding of people in their own contexts. Actually, individuals could use social capital network integration for defining their own behavioural strategy, where social capital network influence decreases with physical distance (Glaeser et al., 2002). In other words, an accurate selection of social capital variables is related to spatial proximity and it has an important influence on travel behavior, because social capital variables generate more local trips (i.e. voluntary activities are locally based). In terms of transport policy, if the local community is social capital oriented, potential public policies to be implemented (including transport initiatives) may differ.
On the other hand, the influence of land use patterns on travel behavior has been subject of many previous studies, but links between international and Spanish perspectives are rarely dealt. Mitchell and Rapkin (1954) wrote one of the first studies to understand the impact of land use patterns on travel behavior, but since 1990 most part of the studies in this field has appeared. Some of them have identified relevant links (Frank and Pivo, 1994; Cervero and Kockelman, 1997; Meurs and Haaijer, 2001; Hong et al., 2013), while others have not found almost any effect (Kitamura et al., 1997; Bagley and Mokhtarian, 2002; Schwanen and Mokhtarian, 2005a). Apart from using different statistical approaches, the different types of land use explanatory variables included in the research are a possible explanation of that controversy. A recent term has being used to define the combination of land use variables and street design on travel behavior studies (e.g. in Ewing and Cervero, 2001; Handy et al., 2011): built environment. It includes all kind of spatial proximity variables (i.e. design of the site, population density, street pattern, land use mix, distribution of commerce, spatial structure and transport supply). However, social capital network influence variables above are not included in the built environment on transport studies. Therefore, there is a real scientific interest in testing the inclusion of this type of variables in a discrete choice-modelling framework (Portney, 2005).

Methodologies
Traditionally, travel behavior has been studied from the utilitarian point of view, focusing on instrumental motivations and ignoring psychological aspects that may play a relevant role. These studies are used to analyze transport demand, based on origin/destination surveys together with Discrete Choice Model (DCM) built on the Random Utility Maximization theory (RUM), which offers an experimental approach to the problem of modeling in a discrete choice context (Mc Fadden, 1974; Mc Fadden, 1981; Train, 2009; Ortúzar and Willumsem, 2011). These models consider that the decision agent behaves rationally, seeking to maximize a stochastic utility function which leads to the introduction of a probabilistic component. Axhausen and Gärling (1992) observe that travel behavior models solely based on the Random Utility Maximization (RUM) theory principle are unable to provide a complete understanding of how people make decisions. Although it has long been recognized that urban travel demand derives from the need and desire to participate in activities, RUM models do not incorporate this fact. Despite the growing interest in activity-based modeling within transportation research, there is a need for more knowledge of the causality relation in order to determine how urban layout influences travel behavior. For example, a more distant shopping center may be chosen because it offers greater variety or better prices (Mokhtarian and Salomon, 2001).

The Structural Equation Modeling (SEM), developed in the 1970s, is a relatively new methodology primarily used in the fields of sociology, biology, psychology, and marketing. It was first applied to travel behavior in 1980 (e.g. Reibstein et al., 1980). In general, SEM is a powerful technique for analysis of causal relationships among endogenous variables, and between endogenous and exogenous variables very useful for identify the interrelationships between traveler attitudes and travel behavior. To studying temporal aspects of activity-travel behavior, several studies of activity duration have also been conducted since the mid-1990s using hazard modeling (Elbers and Ridder, 1982). Finally, for example, some exploratory analysis with the cluster analysis procedure has been used to indicate the ways in which the travel behaviors of some individuals in the different clusters vary (Wittwer, 2014). Overall, it is interesting to observe that common behavioral findings have materialized in the presence of considerable methodological diversity since different perspectives.
Considering that transport planners have been particularly interested in understanding the factors that explain modal choices, it should be noted that since the 1990s a new trend in DCM has emerged in which psychological factors are explicitly incorporated in order to enhance the behavioral of the choice process: hybrid discrete choice models (McFadden, 1986; Ashok et al., 2002; Raveau et al., 2010). These models expand on standard DCM by including attitudes and perceptions as latent variables. The complete model is composed of a group of structural equations describing the latent variables in terms of observable exogenous variables, and a group of measurements relationships linking latent variables to certain observable indicators.

**Panel surveys and GPS technology**

Apart from the advances in methodologies and new variables included on the last travel behavior studies, the treatment of a database coming from a panel survey is increasing (it refers to statistical analysis of panel data which consists of observations made repeatedly of the same sample). Among the advantages that panel data offer are: increased statistical efficiency, possibility of improved prediction, and the ability to observe changes and examine behavioral dynamics (Kitamura, 1990a). The disadvantages emerge from the added biases and costs of panel survey, and the increased complexity involved in the analysis.

The most frequent reason for researchers to design a panel survey is the evaluation of the impact of a change in the transportation system, or a specific transportation planning project, especially when the project involves novel elements (Golob et al., 1997). Indeed, Kitamura (1990) gives an overview of panel studies in transport planning concluding that a panel survey uniquely allow to observe changes in travel behavior and to relate these changes to contributing factors.

Panel surveys can be classified into “long survey panels”, which consist of repeating the same survey (i.e. with the same methodology and design) at “separate times, for example once or twice a year for a certain number of years, or before-and-after an important event; and “short survey panels”, which are multi-day data where repeated measurements on the same sample of units are gathered over a “continuous” period of time (e.g. seven or more successive days), but the survey is not repeated in subsequent years. There are a few examples of panel data mostly gathered for general purposes and, unfortunately, interrupted (Ortúzar et al., 2010):

- The Puget Sound Transportation Panel in Seattle ran 10 waves (Murakami and Watterson, 1989), but with a 2-day travel diary, which raised the question whether two days would be enough for an accurate measurement of changes between subsequent waves of the panel.
- In Canada 2 sophisticated panel surveys about daily activity behaviour were conducted this century. One in Toronto with 2 waves, and one in Quebec with 3 waves. Both panels last 12 months, and involved around 300 respondents with relatively high attrition (Roorda et al., 2005).
- In Adelaide, Australia, approximately 1,000 households participated in a fairly revolutionary panel, with a total of 9 waves (every 4 months), aimed to evaluate voluntarily users' travel behaviour change programmes (Stopher and Swann, 2008).
- In Spain there is evidence of at least two panels. The first had 3 waves and only 143 respondents; it was designed to study season ticket use in the Valencia metropolitan area (Ruiz, 2004). The second is the Madrid-Barcelona Corridor Panel (Ruiz et al., 2008) on the long distance travel. Here a big sample (around 3,000 panellists) was used but it was using a CATI protocol.
In Japan a series of 7 panel surveys were taken between 1987 and 2005. The largest panel had 9 waves and was used to measure the effects on shopping behaviour caused by the opening of a new shopping complex, while the last one had only 2 waves and was used to assess the behavioural changes of registrants of special transport services.

The Dutch Mobility Panel is the largest transportation panel ever realised. It included 10 waves and around 2,000 households. (Van Wissen and Meurs, 1989). The survey was both multi-period and multi-day but it was not driven by a specific purpose. Unfortunately, the willingness to participate was low and up to 47% of attrition was found between waves.

The German panel mobility surveys (Zumkeller, 2009) is the only survey existing today at a national level. Here a 3-year rotating panel was used, where each year about 350 “fresh” households are recruited. On average the attrition rate is around 30%. One limitation is that to reduce burden only residence address was captured in sufficient detail to be geo-coded.

The Mobidrive survey is maybe the most famous short panel dataset. It was a six-week activity diary held in Karlsruhe (Germany) in 1999, which involved 160 households and 360 individuals in the main survey (Axhausen et al., 2002). A similar experience was then repeated in Switzerland (Thurgau, 2003) with the main purpose to detect rhythms of daily life.

The Santiago Panel in Chile (Yáñez et al., 2010b) combine both short and long surveys, but only trip to work were reported and only for a particular day. Individuals were then asked if they made the same trip every day of the week. Moreover, they did not collect all the trips of the day but only asked if the mode selected for the journey-to-work had been influenced by other activities.

Good quality data are essential to estimate reliable demand models. This motivated the choice of using the most advanced technology to measuring and collecting the characteristics of the current trips. Recently the use of technology such as GPS (often integrated in smart phones) has been proposed to track individual movements. This allows measuring with high precision important characteristics of the trip such as origin/destination and all travel times. Moreover, information is available in real time, all the movements are recorded avoiding the typical errors of omitting short trips or overestimating travel times. Using this technology alleviates also the task of the respondents, as they do not have to remember or take notes of all the trips made. On the other hand, although the technology available is very good, there are still some problems such as capturing correctly the signal, especially in urban areas, which requires studying a methodology (a) to improve the measurement precision and the strength of the signal, to avoid errors in the tracks; (b) to increase the number of coordinates (seconds between signals) reducing the charge consumption; (c) to use appropriated algorithms to align typical zigzag erroneous measures; (d) to use map-matching algorithms to match GPS signals with the network implemented in GIS and finally (e) to automate the passage between the information collected into a dataset ready to be used for the mathematical analyses.

So, while the GPS itself gives information regarding time and space, a more “traditional” survey is still needed to gather information about the activities performed and some other characteristics of the individuals. However, technology such as smart phones or interactive web-page interviews will be used for this task, which has great advantages especially in terms of coding the information. Moreover, when this type of interview is associated with the GPS information, there is also a great benefit in terms of quality of information, because individuals can be presented with the tour they made (as recorded by the GPS) and then asked only to tell the activities they were performing at each destination, with whom, the mode used, how much they paid for, and so on. For the people interviewed, this task is much easier than remembering all the characteristics of their trips. Finally, it is also important to mention that using this technology is
substantially cheaper than the standard method of interviewing and later measuring the level of service in the field of urban transport.

The HABIT project

The data used in this thesis comes from a smartphone-based panel survey of the HABIT project (R+D Programme), funded by the “Secretaría de Estado de Ciencia e Innovación”.

Five research centers have participated in the project:

- Centro de Investigación del Transporte (TRANSyT-UPM).
- Grupo de Aplicación de Telecomunicaciones Visuales (G@TV-UPM).
- Grupo Transporte, Infraestructura y Territorio (TIT, UCM).
- Universidad Católica de Chile (UC).
- Danmarks Teckniske Universitet (DTU).

The project had two methodological objectives, directly related to the objectives of this thesis:

- To develop advanced disaggregate models that explicitly account for temporal effects, such as habit and inertia, in mode choice. Such models are crucial to correctly simulate and then forecast users’ responses to the implementation of new policies. Models that fail to account for habit and inertia might severely overestimate demand as well as user benefits due to new policies, leading the administration to take wrong decisions about their implementation. Panels provide the most appropriate data to develop models incorporating temporal effects and, more generally, to study in depth how individuals modify their behaviour as travel characteristics change over time. However, collecting panel data is challenging, and nowadays there are very few panels around the world.
- To build a data panel for the choice of mode; this represented a source of reference not only in Spain but also for the international research community in the field.

Although these aims were mainly methodological, the application to the area of Madrid would allow to derive important results that can be directly used to forecast travel demand and to evaluate the benefits of specific policies that might be implemented in the area. This was of particular relevance for a city like Madrid, that had shown to be highly sensitive to its transport problems and was currently implementing policies that significantly affect individual mobility. It is expected to observe important changes in users’ behaviour and to provide the technical and methodological instruments to correctly measure the dynamic evolution of these changes.

Theoretically, the HABIT project tried to focus the conclusions on two novelty terms in travel behavior studies: habit and inertia. The influence of habit (leading to inertia) in the choice process has been largely discussed in the literature (Gärling and Axhausen, 2003). Although some of this work explicitly accounts for the effect of previous choices (i.e. choices made in previous periods) in the current choice, they all assume that individuals still make a trade-off (i.e. go through a decision process) for any choice in any period. To what extent individuals actually repeat the decision process every time (i.e. every wave of a panel) is a matter of discussion. According to Hoeffler and Ariely (1999), preferences are more likely to be constructed when encountering a new domain, which justifies the specification of the trade-offs to explain part of the choice (inertia being the other part) after a change in the environment (such as in most of the
long-term panel data). The question, “how big should be the change to induce individuals to perform in a decision process?” does not seem to have an answer yet (witness the shock effect in Yáñez et al., 2009).

Regarding to the inertia effect, it tried to use short and long panel data to account for this effect in mode choice models. Incorporating inertia in discrete choice models is at the frontier in demand modelling and requires the integration of psychology within classical microeconomic theory (McFadden and Train, 2000). The last 30 years have witnessed a strong divergence between these two theories. In opposition to the economics tradition that assumes the existence of predefined and stable preferences, psychologist have always argued that preferences are constructed on the spot whenever needed, based on the task and context factors present during choice or preference elicitation methods (Srinivasan, 2001). However, a recent trend among behavioural researchers is to postulate that consumers have inherent preferences (a favoured combination of attributes), and through trial and error they learn what they like; preferences then can change and stabilize over time. However, what translates a first decision into a long-term habit is the mechanism that leads us to repeat that first decision without thinking again about the reasons (“ideal trade-offs”) why it behaves in such a way.

At the same time, this project focused on Madrid because in the last 50 years the city has been experiencing an intense growth accompanied by dispersion in land use. This phenomenon, called urban sprawl, had significantly modified the urban mobility structure, its relation with the territory and the habits of the population. In particular, in terms of mode choice there has been a slight increase in the use of motorized modes (motorized trips per person passed from 1.36 in 1996 to 1.79 in 2004) and the habit to use them might affect the results of some environmentally friendly oriented policies that Madrid decided to implement in the forthcoming years (such as, for example, the Madrid Cycling Plan). It was then important and very timely for Madrid to have the best techniques, methods and models in order to maximise the benefits deriving from the implementation of such policies.

To sum up, while the objective of the HABIT project was to study the temporal effects of habit and inertia using a panel survey to evaluate transport policies, this thesis takes into account not only temporal effects but psychological and built environment factors as well.

Hypotheses
The hypotheses justifying this doctoral thesis can be summarised as follows:

- The success of the implementation of transport and urban policies depends on the capacity to forecast users’ reactions when confronted with changes to the system. Taking wrong decisions has costs too high. To avoid this it is crucial to identify the users’ necessities and their possible responses under the implementation of new policies. Thus demand estimation is a key component of any appropriate urban transport planning framework.
- Most travel demand models to date have been based on cross-sectional data. This data structure fails to correctly ascertain how choices vary over time (both short and long-term variability).
- Models that fail to account for psychological and built environment factors might severely overestimate demand as well as user benefits due to new policies, leading the administration to take wrong decisions about their implementation.
- Panels provide the most appropriate data to develop models incorporating temporal effects but they are complex to gathered. At the same time, the new technology available can provide a valid support to facilitate the collection of the panel. But sperimentation is needed.
Madrid has shown to be highly sensitive to transport problems and is currently implementing policies that significantly affect individual mobility. A city such as Madrid can largely benefit from having a regular panel data, using them to explain how urban and transport characteristics shape mobility, to forecast travel demand and to evaluate the benefits of specific policies that might be implemented in the area.

**Definitions**

The definitions of field-specific terms used in the thesis are:

- **Social capital influence**: “The degree of integration of people with respect to their spatial proximity and social relationship” (Di Ciommo et al., 2014).
- **Psychological latent variables**: “Underlying pro-environmental attitudes in a travel context” (Hess et al., 2013).
- **Built Environment**: “Everything humanly made, arranged, or maintained to fulfill human purposes to mediate the overall environment with results that affect the environmental context” (McClure and Bartuska, 2011).

**Subchapter 1.2 – Objectives and structure of the thesis**

There are a lot of perspectives applied to study demand measures to change urban mobility behavior: econometric and psychological theories, different statistical methodologies, treatment of social or/and land use variables, use of a database coming from a longitudinal or a cross sectional survey, etcetera. Moreover, during the last 40 years tens of thousands case studies on travel behavior has been analyzed using some of the possible combinations of the perspectives above. Therefore, it is very difficult to find the midpoint of the most accurate and complete research in understanding the factors that explain modal choices to minimize negative externalities of car-based trips for a specific case study.

The overall objective of the thesis is to develop a stepped methodology that integrates diverse perspectives to evaluate the willingness to change patterns of urban mobility in Madrid. Madrid is the capital and major financial business center of Spain where a number of transport policies have been introduced since 2010 in the context of the economic crisis: a public transport fare increase of about 25% on average, extension of the time frame for fee-paying on-street parking by one hour, renewal of vehicles and stations in all transport modes, implementation of real-time information systems at most metro/train stops, some bus stops and on-board most metro/train and buses, traffic calming measures in most residential streets (30 km/h or less), etc. This thesis provide a framework on how to integrate some research topics already developed in previous studies and emerging methods on travel behavior, when assessing the introduction of any transport policy. Moreover, this thesis adds new evidences from a Spanish perspective to the research debate on how to change patterns of urban mobility. The data used in this thesis comes from the smartphone-based panel survey (first wave, n=255 respondents and the second, n=190) of the HABiT project.

Figure 1 represents the thesis structure, where are written the four sub-objectives of the thesis following the stepped methodology labeled above. Each one of the four following sub-objectives corresponding with a step proposed at this thesis:

- **Objective I (1st step)**: Analysis of causal relationships between both *objective* and *subjective* personal variables, and travel behavior to capture pro-car and pro-public transport intentions.
Objective II (2nd step): Exploring the potential influence of individual trip characteristics and social capital influence variables on transport mode choice.

Objective III (3rd step): Identifying built environment dimensions on travel behavior from Spanish perspective.

Objective IV (4th step): Exploring the potential influence on transport mode choice of extrinsic characteristics of individual trip using panel data, land use variables using spatial characteristics and social influence variables.

Figure 1. Thesis structure

Chapter 1 - Urban mobility: theoretical background of methods; 1.2: Objectives and structure

1.3: Analysis of the social and land use effects as pro-change factors in travel behavior

Chapter 2 - Smartphone-based panel survey

2.1: Pilot survey: a Revealed Preferences survey of New Stations Users (RP-NSU)

2.2: Smartphone-based panel survey design. Testing smartphone technology

2.3: Data check and descriptive analysis

Chapter 3 - Structural Equation Models

3.1: Analysis of causal relationships between both objective and subjective personal variables, and travel behavior to capture pro-car and pro-public transport intentions

3.2: Involving travel behavior from the perspective of land use. Identifying built environment dimensions from Spanish perspective

Chapter 4 - Discrete Choice Models

4.1: Exploring the potential influence of individual trip characteristics and social capital influence variables on transport mode choice

4.2: Exploring the potential influence on transport mode choice of extrinsic characteristics of individual trip using panel data, land use variables using spatial characteristics and social influence variables

Chapter 5 - Hybrid discrete choice model – Testing the effect of the introduction of transport policy measures to achieve a modal shift. A general framework using objective and subjective personal variables

CONCLUSIONS AND RECOMMENDATIONS
6.1: Discussion of results; 6.2: Policy implications; 6.3: Future researches

The thesis is divided into six chapters. Chapter 1 continues with a brief review of how other studies have included social and land use variables as novelty pro-changes factors in travel behavior (subchapter 1.3) and how it has specified, identified, estimated and tested the two methodologies most used in travel behavior knowledge (subchapter 1.4): DCM and SEM. Then, it is necessary to find a way to evaluate jointly both models: hybrid choice model.

Chapter 2 presents the case study used to reach the thesis’ objectives above mentioned. It starts with the results of a pilot testing through a survey conducted in the new Line 2 metro stations inaugurated in March 2011. Specifically, the pilot examines the treatment of the variables that explain time and cost from a subjective point of view. These results are the basis for the panel survey via smartphone carried out in the
following steps. Chapter 2 is completed with a description of this smartphone-based panel survey design (subchapter 2.1) and some descriptive analysis results (subchapter 2.2).

To reach a first understanding of causal interrelationships between traveler attitudes and travel behavior, Chapter 3 examines the smartphone-based panel survey using a SEM methodology split in two parts: subchapter 3.1 where links between direct and indirect effects on travel behavior (treated as pro-car and pro-public transport intentions) have been explored; and subchapter 3.2 where built environment dimensions are defined using a factor analysis.

The treatment of the smartphone-based panel survey using a DCM methodology belongs to Chapter 4 also split in two parts: subchapter 4.1 incorporates two novelty social influence variables, such as participation in some voluntary activities and to be helped for various tasks (child care, housekeeping, etc.) to test their importance and to add them into the traditional discrete choice-modeling framework; and subchapter 4.2 that using same previous methodology explores the potential influence on transport mode choice of extrinsic characteristics of individual trip in a more general framework -i.e. including panel data (waves 1&2) and land use variables.

Chapter 5 presents results of a general hybrid discrete choice model that takes into account the two previous methodologies, but also the results of the four previous studies. Finally, in Chapter 6 main findings and conclusions are discussed (subchapter 6.1), practical implications are drawn (subchapter 6.2), and future researches provided (subchapter 6.3).

**Subchapter 1.3 - Analysis of the social and land use effects as pro-change factors in travel behavior**

**Social capital influence on travel behavior**

When a new infrastructure is introduced, transport planners are especially interested in defining its pro-modal-shift factors. But it is very difficult to know the real causes of a modal shift. Some studies about voluntary travel behavior change have shown how much travel cost, together with cultural and social values, may influence or even determine traveler’s mode choice (Ampt, 2003). For example, measures for promoting public transport are more effective for new residents and frequent public transport users (Fujii and Taniguchi, 2006). Moreover, attitudes are not necessarily associated with observed behavior i.e., a person who has a positive attitude to reduce car use cannot change to another mode if s/he lacks information about available alternatives to shift (Eagly and Kulesa, 1997).

In the other hand, transport is starting to be recognized as a key component of social policy, particularly in light of a number of recent studies (Turner and Grieco, 2000; Stanley and Vella-Brodrick, 2009), which have highlighted the link between transport and social exclusion, i.e., suggesting that low access to mobility can reduce the opportunity to participate in society (Kenyon et al., 2003).

In order to quantify this social effect on travel behavior, some studies have pointed out that an adequate selection of social influence variables it is necessary to define the social embedding of people in their own contexts (i.e. Di Ciommo, 2003). If the local community is social capital oriented, potential public transport policies, to be implemented may be slightly different (Gray et. al, 2006). Placing individuals in social space is an attractive idea to measure the social effect that however poses a problem when distance-decay function is used to define the intensity of social connections (Hautsch and Klotz, 2003). Carrasco and Miller (2006) define three sets of aspects that influence the propensity to perform social activities as latent
variables: individuals’ personal attributes, social network composition, and information and communication technology interaction with social network members.

DCMs are commonly used in transport planning applications. As discrete choice theory is fundamentally grounded on individual choices, the treatment of the inter-dependence of choices among various decision-makers’ and their integration into social networks remains an outstanding challenge (Dugundji and Walker, 2005). Interestingly, due to the increasing awareness of the influence of social variables in travel behavior, it has now evidence from a wide range of studies that individual outcomes are linked to social interaction factors, especially within social groups defined by geographic proximity (Brooks-Gunn et al., 1993; Glaeser et al., 1992; Bauer and Zimmermann, 1999). In particular, Ostrom (1998) modified the RUM theory to include social interaction variables, such as reciprocity, trust and reputation, and obtained improved results. Brock and Durlauf (2001a, 2001b) and Skrondal and Rabe-Hesketh (2003), have expanded the effort to employ social interaction variables in a structural estimation with binary choice models. Durlauf (2002) looks for the empirics of social capital and explain how a survey could be oriented to detect social capital structure and provide deeper controls for individual heterogeneity.

Therefore, social influence needs to fit in a good way to study the needs of people according to their life cycle stage (Portney, 2005). Páez and Scott (2007) conclude that “a matter for further research would then be to investigate the actual number of contacts that typical individuals maintain in the workplace, as well as in other types of social settings relevant for transport research”.

The spatial dimension on travel behavior
The more recent legislation aiming at stricter mobile sourced emissions control and planning for dramatic decreases in Greenhouse gas emissions emphasizes the need for integrated land use policies with transportation policies. This integration requires understanding of a household residential location change and promotes a move to environmentally friendly behaviors.

Since 1970s the most frequently quoted studies in this respect have provided important conclusions. Through a regression analysis, Hurst (1970) demonstrated that higher rates of vehicle trip generation were found among retail and office land uses compared with storage and industrial usage. Newman and Kenworthy (1989) found a significant negative statistical correlation between residential density and transportation-related energy consumption per capita. Polycentric and dispersed metropolitan areas were found to facilitate shorter commuting times, and differentiation among types of densities turned out to be important (Gordon et al., 1989). In these studies, results for public transportation and automobile were found to be similar.

The impact of density, diversity, accessibility, and percentage of multifamily residential on travel time was study by Ewing et al. (1994). The most relevant conclusion was that households in sprawling suburbs were found to generate around 70% more vehicle hours of travel per person than comparable households in traditional neighborhoods. Also Friedman et al. (1994) distinguished two neighborhood types: standard suburban and neo-traditional neighborhoods. They found that higher total household trip rates and automobile trip rates were found among residents of standard suburban neighborhoods. To find a study that analyzes different neighborhood design, Hess et al. (1999) demonstrated that urban neighborhoods with small blocks and extensive sidewalk systems were found to generate three times more the pedestrian volumes than suburban sites with large blocks and short.
Handy (1996) was among the first to mention the importance of perceptions and attitudes towards land use. Her research revealed that land use is rather a secondary factor in pedestrian choices, although it becomes more important if the walking trip has a destination. Bagley and Mokhatarian (2002) concluded attitudes and lifestyle have a greater impact on travel demand than land use characteristics. But also there are studies point to a higher significance of land use compare to socio-economic and demographic characteristics (Schwanen and Mokhatarian 2005a; b).

In Europe, the SESAME (1999) research project, studying 57 urban agglomerations in France, Germany, Great Britain, the Netherlands, Switzerland, and Spain, pointed to the existence of several important relationships at an aggregate level among land use patterns, travel behavior, and transit supply. These relationships include the following: (i) increasing density is strongly correlated with an increase in the supply of public transport and the presence of rail-based systems; (ii) an increasing concentration of jobs and inhabitants means a higher density of public transport stops; (iii) density was correlated with the share of public transport (positively) and car trips–only drivers (negatively); and (iv) the job concentration in the center was negatively correlated with car use and positively with bicycle use. Other European studies also concluded that land use patterns influence travel behavior (Dargay and Hanly, 2004; Naess, 2005). An exploratory study in this field with a database of Madrid (La Paix et al., 2012) reveals that people living in outskirts areas are likely to multistage tours out of the residence area and the public transport trips decreases with the distance to Center Business District (CBD).

Subchapter 1.4 – Assessment methodology for this thesis

Structural equation models

SEM represents an evolution and a combination of two types of statistical methods: factor analysis and simultaneous equations models (Kaplan, 2000). In SEM, variables can be either exogenous or endogenous (Golob, 2003). These characteristics allow SEM to handle indirect and multiple relationships. Due to these characteristics, SEM is particularly adequate as a tool to model the complex relationships between travel behavior and psychological variables or land use patterns.

The formulation of a SEM with observed variables consist of (Schumacker and Lomax, 2004): a vector of p endogenous variables, a vector x of q exogenous variables, a vector ζ of p disturbances with variance-covariance matrix Ψ, a p×p matrix β that contains the coefficients for the equations relation the endogenous variables, and a q×q matrix Γ that contains the regression coefficients for the equations relating endogenous and exogenous variables. The general equation (no measurement submodels) is:

\[ y = \beta y + \Gamma x + \zeta \]

The model-replicated combined variance-covariance matrix of observed endogenous (p) and exogenous (q) variables, arranged so that the endogenous variables are first, are given by the partitioned (p+q×p+q) matrix (Mulaik et al., 1989):

\[
\sum(\theta) = \begin{bmatrix}
(I - \beta)^{-1}(\Gamma\Phi^{-1} + \Psi)(I - \beta)^{-1} & (I - \beta)^{-1}\Gamma\Phi \\
\Phi\Gamma^{-1}(I - \beta)^{-1} & \Phi
\end{bmatrix}
\]

where Φ is the covariance matrix among exogenous variables. Estimation of SEM models is performed by using the covariance analysis method-method of moments (Golob, 2003). The objective function is to
minimize the differences between the sample variance-covariance matrix \( S \) and the model-replicated matrix \( \sum(\theta) \).

To testing a nested sequence of SEM including previous estimation, Mulaik and Millsap (2000) presented a four-step approach. **Step 1** is to specify an unrestricted measurement model by conducting an Exploratory Factor Analysis (EFA) to determine the number of factors (latent variables) that fit the variance-covariance matrix of the variables observed. **Step 2** involves a Confirmatory Factor Analysis (CFA) that tests hypotheses about certain relationships between indicator variables and latent variables. **Step 3** involves specifying relationships among the latent variables in a SEM approach. Certain relationships among the latent variables are set to zero, so some latent variables are not dependent on other latent variables. If an acceptable fit is achieved for the structural model, the researcher continues to **Step 4**, which tests planned hypotheses on free parameters in the model.

Schumacker and Lomax (2004) formulate the SEM approach and state that most of the model modifications to obtain acceptable data to model fit occur while formulation the measurement models. In fact, a researcher could begin generating the model by using EFA on a data sample to find the number and type of latent variables in a plausible model. Once a plausible model is identified, the same data sample could be used to confirm or test the model, in other words, CFA. EFA (**Step 1**) is even recommended as a precursor to CFA (**Step 2**) when the researcher has no substantive idea underpinning the model.

Schumacker and Lomax (2004) develop the SEM (**Step 3**) in three sub-steps: (i) model ‘specification’, (ii) model ‘identification’, and (iii) model ‘estimation’. Regarding model ‘specification’, SEM does not determine which model to test; it estimates a model’s parameters once that model has been specified a priori by the researcher based on theoretical knowledge. Model specification -i.e. how to define the previous equation (2)- is therefore the hardest part of SEM. The next step is to determine whether the model is ‘identified’, i.e., whether the factor loadings, measurement errors, structure coefficients and specified prediction errors can be estimated. This requires the number of free parameters specified to be estimated in the next sub-step to be less than or equal to the number of distinct values in the sample \( \Psi \) matrix. The direct effects in the SEM model are given by the parameters of the \( \beta \) and \( \Gamma \) matrices, can be interpreted in the same way as regression coefficients, and are the direct influence that one variable has another. For an identified SEM model, the total effects of the exogenous variables on the endogenous variables are given by \((I-\beta)^{-1}\Gamma\) and the total effects of the endogenous variables on one another are given by \((I-\beta)^{-1}I\) (Golob, 2003); they are deducted from the general model expression solved in order to \( y \). Total effects are the sum of both direct and indirect effects. The indirect effects are given by the differences between the total and direct effects. They capture the influence of a variable on another variable through a third mediating variable, thus helping to identify self-defeating policies due to contrary direct and indirect effects.

Unknown parameters can be ‘estimated’ using different estimation methods. The availability of methods depends on the software used. AMOS, in addition to the most commonly used technique of Maximum Likelihood (ML), also includes Unweighted Least Squares (ULS), Generalized Least Squares (GLS), Weighted Least Squares (WLS) and Asymptotically Distribution Free weighted least squares (ADF). ML and GLS assume a multivariate normal distribution of the indicators, which seems unrealistic with empirical data; however simulation studies have shown that ML is fairly robust against violations of this assumption (Boomsma, 2000; Golob, 2003).

The overall goodness of fit indices of a SEM (**Step 4**) are discussed comprehensively in the relevant specialist literature (Mulaik and Millsap, 2000; Schumacker and Lomax, 2004), and most contain similar recommendations. Among others, \( \chi^2/df<2.5, \text{RMSEA}<0.05 \ (<0.08) \text{NFI, GFI, AGFI}>0.95 \ (>0.9), \text{CFI}>0.97 \).
(>0.9) are considered a good model fit, i.e., useful to use. Numbers in parenthesis show acceptable fit values.

Discrete choice models

The analysis of travel behavior is typically disaggregated, meaning that the models represent the choice behavior of individual travelers (Ben-Akiva and Bierlaire, 1999). Ortúzar and Willumsen (2011) indicate that the choice of a transport mode is the key role played by public transport making. DCM is the methodology used to analyze and predict travel choices based on RUM principles. Decision-makers are assumed to select the alternative with the highest utility. For an individual i and for each alternative j the utility function $U_{ij}$ is a function of the observable characteristics $X_{ijq}$ of the alternative (Level Of service, LOS), choice situation (trip purpose, for example) and decision-maker q (income, age, usual place of residence, gender, social integration variables), a set of parameters $\beta_{jq}$ to be estimated, and has associated a random component $\xi_{ij}$:

$$U_{ij} = V_{ij} + \xi_{ij}$$

(3)

where $V_{ij} = \sum q \beta_{jq} X_{ijq}$

The distribution of random component in DCMs is generally assumed to be either normal or Gumbel function (McFadden, 1974). The normal assumption results in the Multinomial Probit (MNP) model (Daganzo, 1979) which allows complete flexibility in the variance structure of the error terms. Nowadays, their use is likely to be limited due to conceptual, computational and statistical problems, including difficulty in interpretation of covariance parameters and forecasting the effects of introducing new alternatives (Horowitz, 1991). If the random component is distributed with a Gumbel function appears the Multinomial Logit (MNL) model (McFadden, 1973), the most widely used discrete choice model. MNL has the advantage of a closed form mathematical structure with simplifies computation in both estimation and prediction. However, it imposes the constraint that the relative probabilities of each pair of alternatives are independent of the presence or characteristics of all other alternatives (Independence of Irrelevant Alternatives, IIA) (Train, 2009). This property implies that the introduction or improvement of any alternative will have the same proportional impact on the probability of all other alternatives. This representation of choice behavior will result in biased estimates and incorrect predictions in cases that violate these conditions. In a MNL the choice probability for each alternative $j$ is then defined as (Ortúzar and Willumsen, 2011):

$$P_{ij} = Prob(\xi_{ik} \leq \xi_{ij} + V_{ij} + V_{ik}) = \frac{\exp(V_{ij})}{\sum_{k=1}^{K} \exp(V_{ik})}$$

(4)

where the scale parameter $\lambda$ is not identifiable (so it ends up estimating coefficients multiplied by it) and $V_{ij}$ is linear function of the coefficients $\beta_{jq}$, but not of the attributes $X_{ijq}$. It includes several interactions among the LOS attributes and decision-makers characteristics q that allow accounting for systematic heterogeneity in the individual preference.

Other DCMs may be derived to relax the three classical restrictions of the MNL model, i.e., random taste variation, unrestricted substitution patterns and correlation of unobserved factors over time (Ortúzar and Willumsen, 2011). For example, Nested Logit (NL) model allows dependence or correlation between the utilities of pairs of alternatives in common groups (McFadden, 1978). From McFadden’s Generalized Extreme Value (GEV) model derive not only MNL and NL, but also the Ordered Generalized Extreme
Value (OGEV) model (Small, 1987), the Paired Combinatorial Logit (PCL) model (Chu, 1989) and the Cross-Nested Logit (CNL) model (Vovsha, 1997).

In particular, the Mixed Logit (ML) is completely general in the above senses and simulation of its choice probability is nowadays computationally simple (Train, 2009). This highly flexible model structure has also the property that the choice probability of any other random utility model can be approximated as closely as desired by an appropriately specified version of the ML (McFadden and Train, 2000).

The ML model is in particular suitable to account for repeated choices by individual decision-makers, such as those made by the respondents in panel survey with repeated choices during a week and between different waves. This new model specification treats the utility coefficients as varying between individuals but remaining through all choice situations for each person (Train, 2009). The ML probably that accounts for panel effect takes the following form:

\[
P_{ij} = \int P_{ij}(\beta) f\left(\frac{\beta}{\theta}\right) d\theta
\]

whereby the conditional choice probabilities are as in (4) but the taste \(\beta\) (namely the alternative specific constants in the case of panel effect) has associated with so called mixing distribution \(f(\beta/\theta)\) over the population (\(\theta\) are known as population parameters). Therefore, the unconditional probability that individual \(i\) will choose alternative \(j\) given a taste of individual-specific coefficients \(\beta\) is given by Ortúzar and Willumsen (2011). The probability in (5) is a multidimensional integral that can be efficiently estimated through simulated maximum likelihood methods.

The ML model is probably the most significant among a number of last innovations in terms of the range of behavior it can accommodate and its overall flexibility. However, the most innovative DCM is the Latent Class Model (LCM). LCM is in some respects a semiparametric variant of the MNL that resembles the ML (Greene and Hensher, 2003). It is somewhat less flexible than the ML model in that it approximates the underlying continuous distribution with a discrete one, however, it does not require the analyst to make specific assumptions about the distributions of parameters across individuals. Thus, each DCM has its limitations and virtues. This thesis uses ML due to his flexibility for a panel database (Train 2009).

Hybrid discrete choice models

The inclusion of subjective elements in DCMs has reemerged as an analysis and discussion topic after losing some of the importance that made it an interesting subject in the 1980s (Train et al., 1987). In the last years hybrid DCMs have been proposed considering not only tangible attributes of the alternatives (classic explanatory variables) but also more intangible elements associated with users’ perceptions and attitudes, expressed through latent variables (LVs) (Ashok et al., 2002). To estimate models with both kinds of variables, two methods have been developed: the sequential approach, in which the LVs are constructed before being entered into the DCM as a further regular variable (Ashok et al., 2002), and the simultaneous approach, in which both processes are done together (Bolduc et al., 2008). The second approach should result in more efficient estimators of the parameters, but it has been used less often because of its greater complexity (Raveau et al., 2010).

Traditionally, in the DCM it is assumed that people \(q\) are rational decision makers maximizing their own utility \(U_i\); the modeler, who is an observer, defines a representative utility \(V_i\) and (since he does not have perfect information) an error term \(\varepsilon_i\) associated with each alternative.
As explained above, the representative utility $V_{ij}$ is a function of the objective attributes $X_{ijq}$ where $q$ refers to a particular attribute (i.e., travel time or fare, as well as socioeconomic characteristics of the individual); if LVs $\eta_{ijl}$ are included, a utility function is obtained:

$$V_{ij} = \sum_q \beta_{ijq} X_{ijq} + \sum_l \theta_{ijl} \eta_{ijl}$$

where $\beta_l$ and $\theta_l$ are parameters to be estimated associated with the tangible attributes and the LVs, respectively. Some studies (Cherchi and Ortúzar, 2011) include several interactions among the attributes $X_{ijq}$ and decision makers’ characteristics that allow accounting for systematic heterogeneity (or taste variations) in the individual preference.

LVs are factors that, although they influence individual behavior and perceptions, cannot be quantified in practice because of their intangibility, since these variables do not have a measurement scale, or because of their intrinsic subjectivity, since different persons may perceive them differently. Identification of LVs requires supplementing a standard preference survey, either revealed or stated, with questions that capture users’ perceptions about some aspects of the alternatives (and the choice context). The answers to these questions generate perception indicators that serve for identifying the LVs. Otherwise, these LVs could not be measured (Galdames et. al, 2011). Moreover, since Cervero and Kockelman (1997) some studies have treated measurable land use variables to define some environmental LVs with satisfactory results (Ewing and Cervero, 2010; De Abreu e Silva et al., 2012, Yoon and Goulias, 2010).

To use LVs, a Multiple-Indicator Multiple-Cause (MIMIC) following a Structural Equation Modeling (SEM) approach model must be estimated, in which the LVs $\eta_{ijl}$ are explained by characteristics from the users through structural equations (7); at the same time, the LVs explain the perception indicators or measurable land use variables $y_{ijp}$, which are observed by the modeler from the survey through measurement equations such as (8):

$$\eta_{ijl} = \sum_r \alpha_{ijr} s_{ijr} + \nu_{ijl}$$

(7)

$$y_{ijp} = \sum_l \gamma_{jpl} \eta_{ijl} + \xi_{ijp}$$

(8)

where: $i$ (individual); $j$ (alternative); $l$ (latent variable); $r$ (explanatory variable); $p$ (indicator).

Since the $\eta_{ijl}$ variables are unknown, the DCM must be estimated jointly with the MIMIC model’s structural and measurement equations. Finally, to characterize the individual’s decisions over the set of available alternatives $A(i)$, binary variables $d_{ij}$ are defined, which take values according to:

$$d_{ij} = \begin{cases} 1 & \text{if } U_{ij} \geq U_{ik} \quad \forall k \in A(i) \\ 0 & \text{in other cases} \end{cases}$$

(9)

Hybrid DCMs can be estimated simultaneously, as well as sequential approach. The former consists of estimating jointly DCM and MIMIC models. While simultaneous is an accurate approach to estimate a hybrid DCM because the estimators from the simultaneous approach are both consistent and efficient (Raveau et al., 2010), the current state of the art does not allow the flexibility to do it correctly in the general case. Moreover, it has been demonstrated that the sequential approach allows for unbiased estimators (Ben-Akiva et al., 2002; Raveau et al., 2010). Therefore, in addition to minor difficulties in its application and interpretation, the sequential approach is chosen by the author to achieve the objectives set in this thesis.
In sequential estimation the problem is treated in two stages, separating the LV and DCM interactions. First, the MIMIC model can be solved to obtain parameter estimators for the equations relating the LVs with the explanatory variables and perception indicators. With these parameters in the structural equation (7), expected values for the LVs of each individual and alternative are obtained. Finally, the LVs can be added to the set of variables of the DCM, and their parameters are estimated with those of the traditional variables.
CHAPTER 2:
SMARTPHONE-BASED PANEL SURVEY
Panel survey design aspects

Over ten years ago, it was established that the most frequent reason that motivates a panel survey on transport studies is the evaluation of a change in the transportation system, or a specific transportation planning project, especially when the project involves novel elements (Kitamura, 1990a). From a statistical viewpoint, a panel survey has the definite advantage to offer more accurate estimates of changes than cross-sectional surveys for the same sample size (Cochran, 2007). Another often cited advantage of panel analysis of multi-day travel behavior (Pas, 1988). Observing travel patterns of individuals and households over several consecutive days, has offered insights into activity scheduling and travel planning. Variability in travel patterns has important policy implications as well. Kitamura (1990) predicted: “panel analysis will be an indispensable tool in many planning and policy contexts as urban areas continue to unfold with changing demographics, socioeconomic, land use, level-of-service, and travel behavior characteristics”. For instance, to achieve the goals of the present thesis it was necessary to design a panel survey.

But also a panel survey has some disadvantages that are necessary to take into account: (i) high costs, in terms of not only time but also money, are typically on of the biggest limitations when building data panels (Yáñez et al., 2010b); (ii) the attrition (loosing respondents) between successive waves; and (iii) the fatigue, i.e., respondents get tired of keeping detailed records of their records of their journeys after a few days (Goulias et. al., 1990).

The concept of panel analysis is referred to the statistical analysis of panel data which consist of observations made repeatedly of the same sample. Therefore, since travel behavior perspective it is necessary to distinguish between (Yáñez et al, 2010b):

- ‘short panel survey’, i.e., multi-day where repeated measurements on the sample of units are gathered over a continuous period of time. One of the first examples was a study of state employees Sacramento and San Francisco to measure the impact of telecommuting on household travel patterns during a three-day survey periods (Kitamura, Goulias and Pendyala, 1990). More recently, it exist the Auxhausen et al.’s (2002 and 2007) studies about the 6 week travel and activity diary data panels collected in Germany and Switzerland, respectively. It is useful for evaluating the variability in travel patterns, but not for evaluating a change in the transportation system.
- ‘long panel survey’, i.e., to repeat the same survey at separate times or different waves; for example once or twice a year for a certain number of years, or before-and-after an important event. Some examples include: (i) a panel study of downtown employees in Honolulu conducted in connection with the evaluation of a demonstration project through two waves that took place before the project and other two during the demonstration project (Golob and Giuliano, 1989); (ii) public transit fare changes underlie the London panel (Terzis, 1988); (iii) the Dutch National Mobility Panel (Golob et al., 1986); (iv) the Pudget Sound Transportation Panel (PSTP) in the United States (Murakami and Watterson, 1990); (v) the German Mobility Panel (Zumkeller et al., 2006); and (vi) Santiago Panel in Chile (Yáñez et al., 2010b). Most part of these experiences describes how the effects of polices could change trends.

For modelling purposes, panel data have been mainly used to study rhythms of life (Axhausen et al., 2002) or activity participation (Jara-Díaz et al., 2008; Susilo and Kitamura, 2005). Researches that have focused only on modal choices have mainly aimed to study how behaviours change as the environment varies (i.e. the supply or the socio-economic characteristics). Golob (1990) employed three waves of data (one year apart) from the Dutch National Mobility Panel to analyse travel behaviour stability. Using structural equation models he found inertial and lagged relationships between income, car ownership, car

Assessment Methodology Applied to Demand Measures to Change Urban Mobility Behavior

Panel survey design aspects

Over ten years ago, it was established that the most frequent reason that motivates a panel survey on transport studies is the evaluation of a change in the transportation system, or a specific transportation planning project, especially when the project involves novel elements (Kitamura, 1990a). From a statistical viewpoint, a panel survey has the definite advantage to offer more accurate estimates of changes than cross-sectional surveys for the same sample size (Cochran, 2007). Another often cited advantage of panel analysis of multi-day travel behavior (Pas, 1988). Observing travel patterns of individuals and households over several consecutive days, has offered insights into activity scheduling and travel planning. Variability in travel patterns has important policy implications as well. Kitamura (1990) predicted: “panel analysis will be an indispensable tool in many planning and policy contexts as urban areas continue to unfold with changing demographics, socioeconomic, land use, level-of-service, and travel behavior characteristics”. For instance, to achieve the goals of the present thesis it was necessary to design a panel survey.

But also a panel survey has some disadvantages that are necessary to take into account: (i) high costs, in terms of not only time but also money, are typically on of the biggest limitations when building data panels (Yáñez et al., 2010b); (ii) the attrition (loosing respondents) between successive waves; and (iii) the fatigue, i.e., respondents get tired of keeping detailed records of their records of their journeys after a few days (Goulias et. al., 1990).

The concept of panel analysis is referred to the statistical analysis of panel data which consist of observations made repeatedly of the same sample. Therefore, since travel behavior perspective it is necessary to distinguish between (Yáñez et al, 2010b):

- ‘short panel survey’, i.e., multi-day where repeated measurements on the sample of units are gathered over a continuous period of time. One of the first examples was a study of state employees Sacramento and San Francisco to measure the impact of telecommuting on household travel patterns during a three-day survey periods (Kitamura, Goulias and Pendyala, 1990). More recently, it exist the Auxhausen et al.’s (2002 and 2007) studies about the 6 week travel and activity diary data panels collected in Germany and Switzerland, respectively. It is useful for evaluating the variability in travel patterns, but not for evaluating a change in the transportation system.
- ‘long panel survey’, i.e., to repeat the same survey at separate times or different waves; for example once or twice a year for a certain number of years, or before-and-after an important event. Some examples include: (i) a panel study of downtown employees in Honolulu conducted in connection with the evaluation of a demonstration project through two waves that took place before the project and other two during the demonstration project (Golob and Giuliano, 1989); (ii) public transit fare changes underlie the London panel (Terzis, 1988); (iii) the Dutch National Mobility Panel (Golob et al., 1986); (iv) the Pudget Sound Transportation Panel (PSTP) in the United States (Murakami and Watterson, 1990); (v) the German Mobility Panel (Zumkeller et al., 2006); and (vi) Santiago Panel in Chile (Yáñez et al., 2010b). Most part of these experiences describes how the effects of polices could change trends.

For modelling purposes, panel data have been mainly used to study rhythms of life (Axhausen et al., 2002) or activity participation (Jara-Díaz et al., 2008; Susilo and Kitamura, 2005). Researches that have focused only on modal choices have mainly aimed to study how behaviours change as the environment varies (i.e. the supply or the socio-economic characteristics). Golob (1990) employed three waves of data (one year apart) from the Dutch National Mobility Panel to analyse travel behaviour stability. Using structural equation models he found inertial and lagged relationships between income, car ownership, car...
mobility and public transport mobility. Bradley (1997), using before and after data collected for 475 commuters, estimated dynamic logit models that account for response lags and state dependence in order to study the effect on mode choice of a new rail commuter line. Simma and Axhausen (2003) used panel data from both Germany and the Netherlands found and that travel commitments (car ownership and public transport season tickets) in one period affect mode usage in the next period. Chatterjee and Ma (2006) used a panel of 4 waves to examine the time-scale of behavioural responses in changes in travel modes where change tends to take longer to occur. Thørgersen (2006) used structural equations modelling and three waves of travel data to study to which extent the current behavior toward public transport is influenced by past behavior, current attitudes and perceived behavioural control.

To evaluate the effects of the transport policies that have been introduced in Madrid during the last four years (five new metro stations, by 25% average public transport fare rise, on-street parking fee rises one hour more, all systems have modernized their vehicles and stations, real-time information at most metro/train stops, some bus stops and on-board most metro/train and buses, most residential streets are traffic-calmed at 30 km/h or less, etc.), it decides to build a ‘short-long’ panel survey based on a sample of a Madrid-worker subpopulation most affected by those recent changes in transport policy: workers in the Regional Health Department within the catchment area of the five new metro stations; and workers in the Universidad Politécnica de Madrid. It discards the most common sampling unit used in transport surveys (i.e. the household survey) to save costs. Moreover, the use of university workers it is useful to control the attrition problem making use of their close proximity to the PhD candidate and the supervisors. But the great novelty that offers this panel survey in comparison to the previous panel surveys is the use of a smartphone application to register each one of the respondent’s trips. Thanks to this smartphone-based panel survey, the fatigue disadvantage is reduced considering that the complete registration of daily trips took about 20 seconds per trip by car or on foot; but also, it provides a real information about the trip using GPS technology.

Characteristics of Madrid’s transport system
Public transport policy in the Madrid Metropolitan Area (MMA) is often deemed as a success. Most of it can be attributed to the “Consorcio Regional de Transportes de Madrid”, created in 1985. Fifteen years ago, the Gross Domestic Product (GDP) per capita in the MMA was about 30% above the average of the E.U. (27,279); this means that the MMA is a relatively wealthy area, even inside the current economic crisis context. However, there are differences in terms of GDP per capita among some cities inside the MMA. Most of the wealthiest municipalities are placed in the northwest whereas the poorest are located in the south east. The GDP per capita of the wealthiest municipalities and neighborhoods is around twice the GDP per capita of the poorest (Vasallo and Perez de Villar, 2008).

According to the 2004 mobility survey for the MMA (EDM-2004), Madrid generated 14.5 million trips per day, which were almost equally distributed on foot (31.1%), private vehicle (34.9%), and public transport (34.0%). Non-discretionary mobility (work and study) represented two thirds of the total mobility. Leaving aside trips on foot, the modal share of public transport represented 49.3% of private vehicle trips. This ratio was substantially higher than it was in other Metropolitan Areas in Europe such as London (39.2%), Paris (29.4%) and Brussels (23.0%) (EMTA, 2007). For these reasons Madrid can be considered as a public transport designed city.

The public transport system in the MMA is made up of four modes. Two of them are typically urban modes (underground and urban buses), and the other two are mostly metropolitan modes (commuter rail and
interurban buses). Beside, since 2007 a new mode of transport (light underground) had been set up in order to increase public transport offer in the outskirts.

The rest of the chapter is organized as follows. Next subchapter presents the results of a pilot testing through a survey (n=1,174) conducts in the new Line 2 metro stations inaugurate in Madrid during March 2011 to evaluate if it is necessary to take into account traveler attitudes and perceptions (of attributes and availability of a mode of transport) on the following smartphone-based survey design. The design and the descriptive analysis of this smartphone-based panel survey are presented in the subchapters 2.2 and 2.3., respectively.

**Subchapter 2.1 – Pilot survey: a Revealed Preferences survey of New Stations Users (RP-NSU)**

As in any survey, the pilot phase is crucial to test the methodology, questionnaire, technology and also to training the interviewers. For the purpose of this thesis, it was planned to test first if it was necessary to include subjective elements (e.g. inertia or shock effect) based on the Madrid-case-study. The results of this pilot survey will be used for the smartphone-based panel survey design.

The Madrid Metro Network Expansion Plan 2007-2011 concluded in March 2011 with the inauguration of several stations on Line 2. The inauguration of any new transport network infrastructure poses a challenge for the assessment of its social impact, and must respond to three questions:
- Does the new transport infrastructure respond to a real mobility need?
- How has the modal shift occurred?
- What is the profile of the user of the new transport lines?

A modeling of Revealed Preference (RP) data based on RUM tends to avoid direct indicators between the modal choice and socioeconomic variables (Mohammadian and Bekhor, 2008; Ben-Akiva et al., 2012), the purpose of travel (Nurul Habib et al., 2009), and the habit or other psychological aspects (Gärling and Axhausen, 2003; Cantillo et al., 2006). But the abovementioned three questions require a response measuring users’ perceptions in modal shift. The psychological Theory of Planned Behavior (TPB) distinguishes three types of beliefs (Ajzen, 1991): (i) normative (cultural context), (ii) behavioral (attitudes), and (iii) control (perceptions). (i) Normative beliefs are related to an individual’s perception of social normative pressures, or the beliefs of relevant others that he or she should or should not perform such behavior. (ii) The attitudes are determined by the total set of accessible behavioral beliefs linking the behavior to various outcomes and other attributes. (iii) Perceived behavioral control is determined by the total set of accessible control beliefs, which are the individual’s beliefs about the presence of factors that may facilitate or impede performance of the behavior. The concept of perceived behavioral control is conceptually related to self-efficacy (Ajzen, 2002).

In travel behavior research there are measured variables for each transport alternative (e.g. time and cost), but users’ self-efficacy must be considered, which is acquired gradually through the development of complex cognitive, social linguistic, and physical skills (Boyd and Vozikis, 1994) as result of experience at different transport alternatives. It is therefore necessary to measure the users’ perception of these LOS variables through subjective variables (Domarchi et al., 2008; Yañez et al., 2010b; Galdames et al., 2011).
Following the TPB theory, recent research reveals factors in the attitudes and perceptions in response to a new transport infrastructure that positively or negatively influence the use of a new transport mode over time (Cantillo et al., 2007; Yañez et al., 2008 and 2009):
- The inertia effect, i.e., the resistance of an individual to change a transport mode for a new one.
- The shock effect, i.e., the novelty the new transport mode represents to the individual.

In this chapter perception variables are introduced to understand changes in travel behavior in response to a new transport network infrastructure, based on data from a Revealed Preferences survey of New Station Users (RP-NSU) on Metro Line 2 including their previous travel mode choice. Since most of the respondents previously made the trip using other transport modes, the main objective of this analysis was to understand which factors (both objective and subjective) influence the modal choice when a new infrastructure becomes available. Specifically, it examines the treatment of the variables that explain time and cost from a subjective point of view. These results form the basis for the mobility panel survey carried out via smartphone for this thesis in the following phase. The panel survey collects data from travelers in different time periods, and includes questions on perceptions and subjective factors.

The second section of this subchapter presents the RP-NSU design and a descriptive analysis of the mobility patterns of the new metro station users. The third section describes the methodology adopted based on RUM and psychological theories for measuring modal shift. Finally, the fourth section gives the most relevant conclusions regarding modal shift when a travel intervention such as a new public transport infrastructure is implemented.

The RP-NSU: Design and Descriptive analysis
A RP-NSU of new metro station users was designed to determine the transport mode used prior to the inauguration, as well as the factors that influenced the shift from the previous to the new transport mode. It should be noted that the sample includes only those people already using the new stations, but provides useful information on Revealed Preferences (RP) which explain the modal shift in response to a new infrastructure, as stated by (Cherchi and Ortúzar, 2002).

The main objective of the RP-NSU was to determine the key factors for the modal shift. More specifically, subjective variables (i.e. perceptions) were introduced in the travelers’ choice. As a preliminary approach, mobility data (travelers/day) was measured for the new metro transport services in the catchment area (around the new stations). One week after the inauguration (March 23, 2011) of the new metro stations – La Almudena, Av. Guadalajara, Alsacia and Las Rosas –, the CRTM carried out a RP-NSU of the users waiting on the metro platform. It is worth noting that RP-NSU has an advantage over other mobility surveys: there is almost no non-response problem, as 90% of the interviewees were simply waiting for their train and had free time to answer. However one disadvantage is that this free time was often very short (1-2 minutes), which constrains the survey design to being very simple and effective. Other objection to RPNSU is that doesn’t provide information from users of other modes of transportation (and potential users of the new metro infrastructures). The number of interviews was 1,174.

The RP-NSU is structured in two sections:
- The first involved the current trip characteristics (itinerary and travel purpose) and the user’s individual socioeconomic indicators (gender and car ownership).
- The second involved trip characteristics before the metro expansion, and the main subjective reason for the modal shift (comfort, perceived speed, perceived cost savings, modernity and safety).
The variation in the total number of users of metro Line 2 was analyzed before and after the expansion, as well as the variation in each of the bus lines running on the same route that were thus directly affected by the new metro stations.

Figure 2 shows the increase in the number of travelers on the whole of Line 2 as a consequence of the inauguration of the four new metro stations. More specifically, average daily travelers increase from 106,268 (before) to 116,622 (after), according to the CRTM database. It should be noted that Line 2 increased from 16 to 20 stations.

![Figure 2. Number of daily travelers on Line 2 (March-April)](image)

Data from the Municipal Bus Company (Empresa Municipal de Transportes - EMT) were used to determine the origin of the new metro users based on the number of daily users of the buses near the Line 2 metro expansion (Figure 3).

![Figure 3. Number of daily travelers on EMT bus lines near to the Line 2 metro expansion (March-April): lines 70, 106, 140, E2](image)

A comparison of the total number of travelers on EMT lines 60, 106, 140 and E2 shows that there is an average decrease of 3,767 travelers/day. Given that the study of Line 2 of the metro revealed that the
number of travelers increased by more than 10,000, it can be concluded that there was a modal shift from both public transport and private cars, as well as a percentage of new travelers who did not previously make the trip. Figures 2 and 3 above show the positive shock effect from a qualitative point of view.

The answers from the 1174 interviewees were analyzed in order to characterize the sample and to identify the explanatory variables to understand the shift in the modal choice before the new infrastructure.

The mobility survey enabled us to classify the interviewee according to gender and car availability. The questions revealed that, despite the RP-NSU restrictions, the sample represents Madrid’s population satisfactorily in these socioeconomic aspects: 56% of the interviewees on Line 2 of the metro were women and 44% men (the Madrid population was 51.76% female and 48.24% male (INE, 2011)); and that 54% had car availability, with approximately the 59% figure for Madrid (Monzón et al., 2013). The following question concerned trip purpose: in the Line 2 stations, 48% of the trips were work related, 15% for study, 4% for medical matters and 33% for leisure/shopping. All the interviewees were metro users at that moment, although two weeks previously they used another mode. Furthermore, they no longer used their former transport mode, or very occasionally, at most in 15% of the cases (Comendador et al., 2012). Thus since this study focuses on modal shift, the answer to the question “How did you use to make this trip”? (Figure 4) is particularly interesting for formulating the DCM later.

**FIGURE 4. Transport mode used by the interviewees before the expansion of metro line 2**

The results above reveal a weak modal shift from private car (14%) and from intermodal trip including car and public transport (5%). RPNSU respondents must answer a key question to evaluate the subjective (or not) reason for changing their transport mode (“Why did you change to the new metro line?”). Possible answers were: comfort, perceived speed, modernity and perceived cost saving. This question provided information on the perception of improvement. Thus, more subjective variables were considered in the subsequent analysis of the DCM: the main reasons to shift transport mode are comfort (43% respondents) and perceived speed (41% respondents).

**Econometric and statistical assumptions for the treatment of a RP-NSU**

The base model used to estimate the transport demand is typically Multinomial Logit (MNL) explained at subchapter 1.4, which is the model that best fits the case study presented in this paper, given the simplicity of the RP-NSU.
MNL models assume that the unknown values $\varepsilon_{ij}$ are independent and identically distributed (iid) with a Gumbel function (McFadden, 1974), in such a way that the probability that the individual $i$ choose the alternative $j$ is given by:

$$P_{ij} = \frac{\exp(V_{ij})}{\sum_{k=1}^{\exp(V_{ik})}}$$

(10)

where $V_{ij}$ is a linear function of the parameters $\beta$ that can be estimated separately and, hence, must be normalized.

On the other hand, if in a further time ($w$) changes occur in some attribute of the alternatives, and it wants to evaluate the probability of the individual changing from alternative $j$ to $r$, then (Cantillo et al., 2007):

$$P_{ir}^w = \text{Prob}\left\{U_{ir}^w - U_{ij}^w \geq I_{ijr}^w \text{ and } U_{ir}^w - U_{ik}^w \geq I_{ijr}^w - I_{ijk}^w \forall k \neq j, k \neq r\right\}$$

(11)

where $I_{ijr}^w$ is the inertia i.e., or the resistance that the individual offers to changing from alternative $j$ to $r$.

As previously stated, the definition of the initial utility function can be restructured depending on whether the temporal wave $w$ for the individual $i$ is:

$$\tilde{U}_{ijr}^w = U_{ir}^w - I_{ijr}^w$$

(12)

where $I_{ijr}^w = f(\text{ijr})$; i.e., that the inertia for the individual $i$ of changing from alternative $j$ to $r$ depends on the utility variation between alternatives in a former temporal phase ($w-1$).

Following the introductory discussion, this study aims to understand the modal shift using a RP-NSU survey. There are two significant differences with a traditional RP survey: (a) this survey does not capture utility differences in mode choice between two waves because the PhD candidate do not have the “before” of the metro trip, since it did not exist; and (b) this survey includes only those people already using the new stations, so it does not have information on other mode choices after the inauguration of the new metro stations. The first difference above makes impossible to capture shock effect labeled for Cantillo et al. (2007) and Yáñez et al. (2008 and 2009) -as well as inertia effect-, since it would be a function of the difference between the utility of the metro evaluated at the current wave $w$, and its utility evaluated at the previous wave $w-1$ (where it did not exist).

To attach a formulation, it can consider a previous wave ($w-1$) in which the pre-opening alternative ($j$) is bus, bus combination, bus/metro combination, car, car/public transport or public transport/on foot. After the inauguration of the new metro stations, it is in a new wave ($w$) in which the only alternative that can be analyzed is the metro.

The rest of the assumptions supporting the formulation are:

- Respondents are utility maximizers, as in RUM theory.
- RP-NSU information does not present panel correlation.
- Utility function measures the degree of satisfaction that people derive from their choices. The utility function of the metro choice depending on the temporal wave $w$ for the individual $i$ is:

$$\tilde{U}_{ij\text{metro}}^w = U_{ij\text{metro}}^w - I_{ij\text{metro}}^w$$

(13)
- The utility function of the metro based on RUM that does not depend on the temporal effect ($U_{\text{metro}}^w$) is considered a constant $K$, since in this RP-NSU the probability that the individual $i$ chooses the metro alternative is almost 100% ($P_{\text{metro}}^w \approx 1$).
- The inertia effect is a function of the previous ($w-1$) valuation of the options (following the approach of Cantillo et al., 2007):

$$I_{ij\text{metro}}^w = f_i (V_{ij}^{w-1} - V_{\text{metro}}^{w-1}) = (\beta_j^w + \cdots + \beta_{ELSE} \cdot SE_i) \cdot (V_{ij}^{w-1} - V_{\text{metro}}^{w-1})$$ (14)

- Given the fact that the metro alternative did not exist before the inauguration of the new metro stations, the measurable part of the pre-opening utility function of the metro ($V_{\text{metro}}^{w-1}$) is considered null.
- A negative inertia effect is the “typical” inertia effect in the absence of changes based on psychological theories (Gal, 2006). This effect is related to the so-called loss aversion by Kahnemann and Tversky (1979); and recent applications in travel behavior research (van der Kaa, 2010; Avineri and Waygood, 2013). In the case of RP-NSU, it is in an “atypical” situation because people choose the metro because the pre-opening alternative caused them a degree of disutility (dissatisfaction). Therefore, PhD candidate assumes a positive inertia effect-modifying its sign in previous equation (6) - that will indicate the preference for changing to the new metro infrastructure.

Summarizing these assumptions, if it has a RPNSU and it wants to measure the utility of a new transport infrastructure (Metro in this case) it must take into account that (i) this depends on a positive inertia effect of each pre-opening alternative disutility ($-V_{ij}^{w-1}$), and (ii) when estimating the utility function of each pre-opening alternative as in RUM theory, the higher the probability of choosing a pre-opening alternative $j$, the higher its positive inertia effect on the metro ($I_{ij\text{metro}}^w$):

$$\bar{U}_{ij\text{metro}}^w = K - f_i (-V_{ij}^{w-1}) = K + f_i (V_{ij}^{w-1}) = K + I_{ij\text{metro}}^w$$ (15)

And as a simplified form for the binomial logit model (BNL), the conditional probability of shifting to the new mode can be written as:

$$P(\text{shift from } j \text{ to METRO}|X_i) = \left(\exp(K + I_{ij\text{metro}}^w)\right) \cdot \left(1 + \exp(K + I_{ij\text{metro}}^w)\right)^{-1} = \exp(K) \cdot \left(\exp(K) + \exp(-I_{ij\text{metro}}^w)\right)^{-1}$$ (16)

Therefore if it estimates a positive $\beta_j$ coefficient associated to a specific $X_i$ variable (e.g. number of children) of the measurable part of the pre-opening utility function of alternative $j$ ($V_{ij}^{w-1}$), an increase in the value of this variable will produce a decrease in the above denominator and an increase in the probability of shifting from $j$ alternative to the metro.

Model definition and results
The results of the RPNSU allow us to know for each individual $i$ trip:

- Pre-opening alternative ($j$): when answering the question “How did you use to make this trip? (bus, bus combination, bus/metro, car, car/public transport, public transport/on foot)”.
- Travel time ($TIME_{ij}$): if it knows both the origin and destination stations, as well as the different modes used in cases of transport combinations, the travel time can be measured thanks to several web apps: Google Maps (©2013 Google), EMT Navega por Madrid (©2010 Empresa Municipal de Transportes de Madrid, S.A.) and Viaja en METRO (©2013 METRO Madrid).

- Travel cost ($COST_{ij}$): an average cost of €0.90 has been estimated for the pre-opening public transport users, while the cost of traveling by car was indirectly computed from trip distance and fuel consumption data using other references and data bases (Di Ciommo et al., 2010; Google Maps, ©2013 Google).

Information on other explanatory variables complementary to the alternatives has been also included ($z_{ij}$):
- Trip purpose: work, study, doctor’s visit, leisure, shopping, other; although it has distinguished obligatory trips: commute trip ($COMMUTE_{ij}$).

- According to the answer to the question “Why have you changed to the new metro line? (comfort, perceived speed, modernity, and perceived cost saving)”; and seeking also to explain the time and cost from a subjective point of view, it was selected two dummy variables: speed perception ($SPEED\_P_{ij}$) and cost saving perception ($COST\_P_{ij}$). The former refers to the speed perception of the pre-opening alternative, and the latter concerns cost saving perception (the search for a cheaper alternative).

The following variables have been identified regarding the socioeconomic characteristics ($SE_i$) of the individuals:
- Gender ($SEX_i$).
- Car ownership ($CO_i$).

The objective is to measure the utility from a previous alternative ($j$) as:

$$V_{ij} = \beta_0 + \beta_C COST_{ij} + \beta_T TIME_{ij} + \beta_SE_i SE_i + \beta_z z_{ij} \tag{17}$$

The DCM was estimated using the language and environment for statistical computing R (2012). The values of the explanatory variables (time and cost) were directly included in the model, as indicated in the previous formulation, whilst the rest of the variables ($SE_i, z_{ij}$) were incorporated as dummy variables. Table 1 summarizes the values adopted for this type of variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Abbrev.</th>
<th>Value 1</th>
<th>Value 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>SEX_i</td>
<td>Man</td>
<td>Woman</td>
</tr>
<tr>
<td>Car ownership</td>
<td>CO_i</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Commute trip</td>
<td>COMMUTE_i</td>
<td>Trip purpose:</td>
<td>Trip purpose:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>work or study</td>
<td>doctor’s visit, shopping, other</td>
</tr>
<tr>
<td>Speed perception</td>
<td>SPEED_P</td>
<td>Reason to change:</td>
<td>Reason to change:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>perceived speed</td>
<td>comfort,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>perceived cost saving, other</td>
</tr>
<tr>
<td>Cost saving perception</td>
<td>COST_P</td>
<td>Reason to change:</td>
<td>Reason to change:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>perceived cost saving</td>
<td>comfort,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>perceived speed, other</td>
</tr>
</tbody>
</table>

A key question for the model specification is the number of alternatives $j$ to be considered prior to the expansion of the metro lines. First, a multinomial logit (MNL) model was considered with the following groups of alternatives depending on the answer given to the question “How did you use to make this trip?”.
Assessment Methodology Applied to Demand Measures to Change Urban Mobility Behavior

(a) public transport: bus, bus combination, bus/metro combination and public transport/on foot; (b) car/public transport; (c) car. The results showed correlations between the results of alternatives (b) and (c); and are the reason a binomial logit (BNL) was finally used, in order to distinguish the trips in which the car was used: (a) versus (b) & (c).

The whole choice test consisted of two alternatives: (1) public transport (PT) and (2) private transport (CAR). It then also checked that the set of choices was mutually exclusive, collectively exhaustive and finite (Train, 2009).

The chosen modes are characterized by means of the estimated LOS and other explanatory variables described above. The most important variables for the purpose of this article are those that reflect, from a subjective point of view, the disutility that respondents attribute to the pre-opening alternative: speed perception ($SPEED_{P_{ij}}$) and cost saving perception ($COST_{P_{ij}}$). This is done using dummy variables and was designed to ascertain whether better models can be achieved when using non-instrumental attributes.

Two different models are estimated to validate the methodology including psychological variables on a DCM based on the RUM theory (“individuals are utility maximizers”): the first without subjective variables (BNL1); and the other adding $SPEED_{P_{ij}}$ and $COST_{P_{ij}}$ in order to increase statistical significance and explanatory power (BNL2). The performance of models with different numbers of parameters can be compared using criteria based on Bayesian theory. The Akaike Bayesian Information criterion (variously abbreviated as AIC) compares maximum likelihood estimation goodness-of-fit and the dimensionality (parsimony) of each model (Akaike, 1974 and 1987). The model that yields the smallest value of AIC is considered the best.

Table 2 shows the results of the BNL model, taking the CAR alternative as a reference. The model output provides asymptotic t-test values for estimated coefficients, and the final value of the Log-likelihood, $\rho^2$ and AIC. The t-test shows whether the estimated coefficient for each variable makes a significant contribution to the prediction of the outcome: utility for pre-opening public transport users.

Table 2. Results from BNL models with statistically significance $\beta$ and 95% confidence interval for $\text{Exp}(\beta)$. N= 1174 respondents

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>BNL1</th>
<th></th>
<th></th>
<th>BNL2</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta$ (t-test)</td>
<td>Lower</td>
<td>$\text{Exp} (\beta)$</td>
<td>Upper</td>
<td>$\beta$ (t-test)</td>
<td>Lower</td>
</tr>
<tr>
<td>PT constant</td>
<td>2.642 (3.4)</td>
<td></td>
<td></td>
<td></td>
<td>4.427 (17.3)</td>
<td></td>
</tr>
<tr>
<td>TIME</td>
<td>0.236 (10.4)</td>
<td>1.314</td>
<td>1.266</td>
<td>1.219</td>
<td>0.201 (8.8)</td>
<td>1.178</td>
</tr>
<tr>
<td>COST</td>
<td>-4.192 (-5.4)</td>
<td>0.054</td>
<td>0.015</td>
<td>0.004</td>
<td>-4.739 (-5.2)</td>
<td>0.002</td>
</tr>
<tr>
<td>SEX</td>
<td>-0.634 (-2.1)</td>
<td>0.321</td>
<td>0.530</td>
<td>0.877</td>
<td>-0.573 (-1.8)</td>
<td>0.301</td>
</tr>
<tr>
<td>CO</td>
<td>-2.560 (-6.5)</td>
<td>0.040</td>
<td>0.077</td>
<td>0.148</td>
<td>-3.139 (-5.4)</td>
<td>0.017</td>
</tr>
<tr>
<td>COMMUTE</td>
<td>-0.168 (-0.5)</td>
<td>0.512</td>
<td>0.846</td>
<td>1.397</td>
<td>-0.646 (-2.1)</td>
<td>0.289</td>
</tr>
<tr>
<td>SPEED_P</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.021 (2.8)</td>
<td>1.509</td>
</tr>
<tr>
<td>COST_P</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-6.8 (-4.5)</td>
<td>0.000</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-141.611</td>
<td></td>
<td></td>
<td></td>
<td>-107.987</td>
<td></td>
</tr>
<tr>
<td>$\rho^2$</td>
<td>0.65</td>
<td></td>
<td></td>
<td></td>
<td>0.73</td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>295.222</td>
<td></td>
<td></td>
<td></td>
<td>231.976</td>
<td></td>
</tr>
</tbody>
</table>

NOTE: two-tailed probability values: *p>0.1; *0.05<p<0.1. Rest of the coefficients: p<0.05.

By including subjective variables (BNL2) it can be seen that the log-likelihood, $\rho^2$, has improved, and the AIC values and both dummy variables for measuring perception are statistically significant. Obviously, this shows that the observed pre-opening alternative choice – used to measure the positive inertia in the
change to the metro—is not only ruled by the RUM theory, but also by a strong perception component, developed without the mediation of any cognitive processes during the choice. Perception variables increase choice probability for PT or CAR, with a non-linear influence on the utility function.

The results in Table 2 show that a respondent concerned with the SPEED_P of his pre-opening alternative is more likely to be a previous PT user, and a previous PT user is more likely to change to the metro due to this perception and not to others such as comfort, economy or security. This subjective explanation matches the objective interpretation of the positive value of the TIME coefficient. However, since the SPEED_P variable reduces the relative importance and the statistical significance of the TIME variable, it is probable that the SPEED_P perception was hidden behind the TIME variable in the BNL1 model. The same occurs between the COST_P and COST variables but with a positive effect on the CAR utility function.

As a rule, the traditional DCM results show negative TIME and COST coefficients when the CAR alternative is the reference, but the longer trips recorded in the RP-NSU—“atypical” situation, as indicated in the previous section—are more likely to be associated to pre-opening PT users, because long trips are more likely to characterize CAR users who have not changed to the metro (RP-NSU doesn’t take into account these users). However, as shown above from a subjective and objective point of view, a reason for modal shift by car users is the economic aspect. Surely the economic crisis in Spain might be also affecting this last result.

An analysis of the SEX variable reveals that there is a higher likelihood of male car users to change to the metro than female, probably because women usually take on greater household responsibilities (childcare, etc.) making it difficult to make a modal shift. Finally, commute trips (work and studies) reduce the resistance to change (negative inertia effect) between the car and the new metro infrastructures more than any other trip purpose (doctor’s visit, leisure, shopping, etc.). This is because it is easier to make a modal shift for a habitual trip than for non-habitual trips.

Finally, it can be said that the incorporation of perception questions in a RPNSU for application in a DCM model—even with a simplified survey like the one used here—improves the fitness and statistical quality of a BNL model that directly measures modal shift through the utility functions of the pre-opening alternatives.

The importance of subjectivity on travel behavior
Data gathered from a Revealed Preferences pilot survey of the New Station Users (RP-NSU) allows the explanation of the modal shift by considering inertia effect (Cherchi and Manca, 2011) as one variable explaining the individual’s utility function regarding the modal choice.

Thus, the descriptive analysis of the data from the RP-NSU carried out in the metro (Figure 1), together with the data for the variation in the number of EMT passengers (Figure 2) suggests that there has been a positive novelty effect regarding the change in transport mode, as there was an immediate transfer of former EMT users to the metro from the inauguration. This mobility never increases over time but remains at a middle term (Comendador et al., 2012).

The inertia modal shift is analyzed using a BNL model based on RUM theory, including subjective variables that improve the fitness and statistical significance of the model. Hence by simplifying the
general theory of the inertia effect, a positive relationship is found between the pre-opening disutility alternative (positive inertia) and the likelihood of changing to a new and unknown mode.

The results of the model point to the following policy recommendations to be considered across the rest of the thesis:

- One of the main reasons why modal shift from car is so difficult to achieve, is the perceived travel time. An urban trip by car is always perceived as taking less time than on public transport. Therefore, by way of example, in the case of new metro infrastructures, the speed of the trains should not be the only point to be stressed.

- Pre-opening car users’ perceived cost savings from a short-medium trip by metro shows less resistance to the modal shift from private to public transport. Nevertheless, since car users only perceive their fuel costs –not captured by our RP-NSU–, a policy question worth addressing would be to make them aware of the real nature of the costs of car use (depreciation, insurance, taxes, maintenance and inspections).

- Commute trips (work and studies) are more likely to produce a modal shift than other trip purposes. This paves the way for green transport plans (packages of practical measures to reduce car use for journeys to and from work and for business travel), as an effective and sustainable response to increase public travel demand on routes with limited or non-existent public transport services (López-Lambas and Aparicio, 2004).

**Subchapter 2.2 – Smartphone-based panel survey design**

**General aspects**

The data used in this thesis comes from a smartphone-based panel survey conducted in the HABIT project (Habit and inertia in mode choice behavior: a data panel for Madrid). This panel survey has features of both ‘short’ and ‘long’ panel surveys used for other authors. ‘Short’ because a smartphone with a survey application was given to the respondents to capture all the trips during a working week (five days). In fact, Cherchi et al. (2009) showed that if the interest of the panel was in modelling mode choice, the ideal length of the panel could be just a week, provided the sample is not too small. And ‘long’, because the smartphones –with a survey application- were distributed at two different points in time (2011-2012). At the beginning, the PhD candidate and the supervisors had conceived the panel with three waves: one before the new metro stations inaugurated in Madrid during March 2011; and the second and third waves during the sixth (September 2011) and eighteenth months (September 2012), respectively, after the inauguration. However, due to administrative troubles, it was decided to start the first wave on September 2011 and the second wave one year after in order to study the willingness to change patterns of urban mobility with the introduction of transport policy measures (not only new public transport infrastructures, but also those mentioned in the introduction of this chapter).

To compute the first wave, during fall 2011 and winter 2012, two groups were given a smartphone with the panel-survey application for one week: (1) 91 workers from the Regional Health Department within the catchment area of the new metro stations; and (2) 164 workers from the Universidad Politécnica de Madrid- making use of their close proximity to the PhD candidate-, thus easily producing a random sample of 5774 workers (2011 census data). As it had been indicated previously, it discarded the most common sampling unit used in transport surveys (i.e. the household), and based the panel survey on a sample of a worker subpopulation. In fact, the treatment of this subpopulation and the smartphone technology are
useful to reduce the effect of the three most important disadvantages in a panel survey: high cost, attrition and fatigue. The second wave took place during fall 2012 and winter 2013 with a total of 190 respondents.

Before conducting the panel survey, three focus groups were organized with residents of the new metro stations’ catchment areas; these allowed improving the questionnaire design, to understand the perception of the new metro lines and to determine the profile definitions of the residents and workers to be recruited (Aizer and Curie, 2002). The focus groups were also used to design the incentives to attract individuals to participate in the survey.

Sample size determination
Following the approach used in the Santiago Panel (Yáñez et al., 2010b), the simple size (n) of the smartphone-based panel survey is computed using the formula (Smith, 1979):

\[ n = \frac{CV^2 \cdot Z_\alpha^2}{E^2} \]  

being CV the coefficient of variation of the variable under thesis, E its level of accuracy proportion and \( Z_\alpha^2 \) the standard normal value for the \( \alpha \) confidence level. As it wants to measure the willingness to change patterns of urban mobility (following DCM methodology), it is necessary to ensure the presence of every mode. Monzón et al. (2013) fixes the Madrid’s modal distribution as: 45% (car), 40% (public transport) and 12% (walk and bike). It determines the CV value according to:

\[ CV = \frac{\text{st. deviation}}{\text{mean}} = \sqrt{p(1-p)} / p; \ p = 0.12 \]  

Then, assuming a level of accuracy equal to 25% and a confidence level of 90% (\( Z_\alpha^2 = 1.282 \)), the sample size should be:

\[ n = \frac{((\sqrt{0.12(1-0.12) / 0.12}) \cdot 1.282)^2}{0.25^2} = 193 \]  

Moreover, to justify the sample size for the SEM applications used in this thesis, although sample size is important in factor analysis (first step in a SEM methodology), there are varying opinions and several guiding rules of thumb are cited in the literature. Hair et al. (1995) suggested that sample sizes should be 100 or greater. A number of textbooks cite the work of Comrey and Lee (1973) in their guide to sample sizes: 100 as poor, 200 as fair, 300 as good, 500 as very good, and 1000 or more as excellent.

Summarizing, the sample size used in two waves (255 and 190, respectively) is reasonable to justify the objectives of this thesis.

The smartphone-based panel survey and its evolution
As stated above (see subchapter 1.1.), five research centers participated in the smartphone-based panel survey design. A so huge group was necessary since it was the first experience in a panel survey that collected these three aspects: (i) the development of a smartphone application to register each one of the respondent’s trips (by G@TV-UPM); (ii) the definition of the GIS information necessary to establish some
urban environmental variables; and (iii) the benefits of the Santiago Panel in Chile (UC and DTU) as the last most famous panel survey developed.

The smartphone-based survey (Figure 5a) for the first wave (September 2011) had two main phases. The first consisted of a face-to-face interview to gather personal data on the respondent. As in the more typical cross-sectional surveys, it considered the variables age, sex, education level, income, number of working hours per week, number and type of cars, number of bicycles and possession of driver’s license, etcetera. Moreover, following the results of the pilot survey, respondents also answered an attitudinal and perception questionnaire with Likert scales covering comfort, security, and accessibility topics about Madrid’s public transport; this interview was also used to explain the context and objectives of the survey (Figure 5b). In the second phase, they were given the smartphone and asked to register the daily trips they made during the five workdays (Monday to Friday) (Figure 5c). The variables considered were mode used, departure and arrival times, transfer times (for intermodal or mixed modes), cost (fare for public modes; parking and toll charges for private modes), and number and location of mode interchanges.

**Figure 5. An overview of the smartphone application**

(a) Starting the application  
(b) Answering attitudinal and perception questionnaire  
(c) Recording a trip (LOS and GPS information; and other characteristics)

The trips were monitored in real time, and respondents were contacted at the end of the day to correct or clarify the information. The participants were also given a chart to manually register any trips that were not recorded on the smartphone. The complete registration of daily trips took about 20 seconds per trip by car or on foot, and one minute for a journey by public transport. At the end of the trip, the data were automatically sent to a server accessible by the PhD candidate. Updating the first wave questionnaires (face-to-face interview and level of service measures) was an interactive procedure between the previous five researcher centers that involved three steps (Goulias et al., 1990): (i) using three focus groups to identify questionnaire items that may had cause inaccurate responses or technical problems with the application, as it explained in the beginning of this subchapter; (ii) with this information, changes to the layout, presentation, wording and contents were made; and (iii) the changes were evaluated between the people of the HABIT project, and the process was iterated by going back to either previous steps.

For the second wave (September 2012) a few changes were introduced with the experience of the first wave, but without producing any substantial change. The complete questionnaire is presented in the ANNEX.

When revealed preference data are gathered, it is crucial to clearly define which modes individuals have available for each trip. DCMs are based on the relative evaluation of the chosen alternative compared to all the other alternatives available for each individual. Then the same quality required for the
characteristics of the chosen alternatives must be achieved for the characteristics of the non-chosen, but available, alternatives. To measure this information was crucial to carefully process the data gathered in order to check for all those aspects that might constrain the availability of some alternative, such as tours made during the whole day, interactions among family members and/or friends, collecting materials or picking up children at school and so on. This diagnostic was needed to carefully define the choice set for each trip of each individual.

Having done that, the characteristics of the non-chosen modes needed to be quantitative measured using times scheduled from the public transport companies (EMT Navega por Madrid ©2010; Viaja en Metro, ©2013) and with direct measurements in the network with instrumented vehicles (©2013 Google).

Although initial quality is very important in any type of survey, panels present an additional requirement: it is necessary to maintain quality over time. In that respect the main problem is how to avoid attrition (Kitamura, 1990b); for this various effective maintenance methodologies were applied, such as:

- **Providing incentives.** This is useful in panels as it is necessary to maintain respondents motivated across waves (Yáñez et al., 2010b). Some authors have reported using cash or gifts as incentives, but others raffled different prizes. Since the HABIT Project financing, both incentives are used; but the latter had better acceptance (Figure 6).

**Figure 6. Incentives used for maintaining respondents**

- **Maintaining contact.** The idea was to keep respondents interested in the study. Thus, it was sent short e-mail reminders, summary reports and letters of gratitude to some participants between waves (Figure 7).
Conducting face-to-face interviews. Even though this is the most expensive way to gather data, it was decided to use it because it provided the best alternative in terms of answer quality and response rates (Tourangeau et al. 1997).

Using the same interviewer for the same respondents, as other panels have reported managing to attain lower attrition rates with this measure (Van Wissen and Meurs 1989; Hensher 1985). Thus, the PhD candidate was personally responsible for delivering and picking all the smartphones to the respondents.

Applying these maintenance methodologies it was possible to retain the 75% respondents from the first (255) to the second wave (190). As mentioned before, some incentives were provided as a means to control attrition; furthermore, different types of incentives were tested to know which one could have a higher impact. Unfortunately, it was not possible to extract conclusive recommendations for future panels, as attrition figures were so low.

**Subchapter 2.3 – Data check and descriptive analysis**

After a brief discussion on general aspects about the design of the survey and with the benefit of the experience of the pilot survey, it is necessary to provide descriptive and quantitative analyses of the current characteristics of mobility in Madrid, and to highlight the first causal relationship between mobility, urban form and socio-economic characteristics. This phase is crucial to set the background of the thesis, identify the area of application, define the sample and its characteristics, and to acquire a first picture of the attributes that could be relevant in modelling mode choice.
The analyses performed in this point are mainly based on the information provided by the *Encuesta Domiciliaria de Movilidad* (Mobility Household Survey - EDM) conducted by the Consorcio de Transporte de Madrid (CTM) in 2004, which is, despite the date, the best available material to reach an appropriate knowledge of the current situation. Data from INE (2011) and Monzón et al. (2013) are used to compare characteristics of the worker population of Madrid with a sample. To understand the urban model, Geographical Information Systems are used for the incorporation and analysis of available cartography for the Community of Madrid: land use, land survey and planning department information for each municipality contained in the MadPlan application. The transport systems analysis will be likewise supported by the GIS, using the information georeferenced from the Networks included in the *Mobility GIS* of the CTM.

All the available data is analyzed in depth in order to provide the following features for the area of Madrid:

1. Analysis of the urban structure in terms of location of major activities, density of population, density of workers for type of work, opening and closure times of activities, type of urbanization and so on.
2. Analysis of the socio-economic characteristics of the population such as age, sex, family size, number of children, income, occupation and so on.
3. Analysis of the present characteristics of the transport system in terms of available modes and their main characteristics such as times (in vehicle, waiting, walking, parking, transfers and so on), costs, interchanges, frequency and so on.

Firstly, although the panel sample included only workers from the Universidad Politécnica de Madrid and Regional Health Department, their household locations are satisfactory widespread over the city of Madrid. The changes of home location are limited mainly by the proximity of the two waves. In fact it observed that around 3% of respondents changed location between waves. These home location changes did not generate repercussions in terms of attrition rates. On the other hand, some work changes also took place between waves and these are responsible for part of the attrition (as it uses work-place interviews). Fortunately, the proportion of work changes has been low as well (2%).

Secondly, regarding the traditional socioeconomic characteristics, Table 3 shows that while the average age increased slightly between waves and the number of family members decreased slightly, the average income presented more variation. Also Table 3 shows that despite the restrictions for choosing the respondents, the sample well represents the Madrid worker population (INE, 2011; Monzón et al., 2013) in these aspects. The reason for the income increase between the first and the second wave is that many of those 65 respondents that did not repeat the survey are people with eventual and precarious employment contracts.

### Table 3. Sample socioeconomic characteristics

<table>
<thead>
<tr>
<th>Variables</th>
<th>Natural Travelling Working Population in Madrid Metropolitan Area*</th>
<th>Wave 1 (n=255)</th>
<th>Wave 2 (n=190)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Average St. Deviation</td>
<td>Average St. Deviation</td>
</tr>
<tr>
<td>Male (%)</td>
<td>51</td>
<td>52 9.2</td>
<td>51 11.6</td>
</tr>
<tr>
<td>Age</td>
<td>40</td>
<td>43 9.2</td>
<td>44 11.6</td>
</tr>
<tr>
<td>Income (€)</td>
<td>2500</td>
<td>2100 410</td>
<td>2220 320</td>
</tr>
<tr>
<td># family members</td>
<td>2.7</td>
<td>3.1 0.5</td>
<td>3.0 0.3</td>
</tr>
</tbody>
</table>

*Source: INE (2011) and Monzón et al. (2013)
Furthermore, a new labor law that increased the number of working hours/week was introduced by the Spanish government between the first two waves (February 2012) and, at the same time, some of the respondents increased their salary, while others suffered a decrease mainly due to the reduction in their Christmas bonus. And thirdly, regarding mobility characteristics there are more differences between waves, as it can see at Table 4.

**Table 4. Daily mobility characteristics**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Mobility Working Population in Madrid Metropolitan Area</th>
<th>Wave 1 (n=255)</th>
<th>Wave 2 (n=190)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Average St. Deviation</td>
<td>Average St. Deviation</td>
</tr>
<tr>
<td># trips</td>
<td>2.6</td>
<td>2.4 0.3</td>
<td>2.1 0.4</td>
</tr>
<tr>
<td>Travel time (min)</td>
<td>28.6</td>
<td>32.7 5.8</td>
<td>30.5 6.1</td>
</tr>
<tr>
<td>Commuting distance (km)</td>
<td>6.0</td>
<td>7.9 3.7</td>
<td>6.5 2.6</td>
</tr>
<tr>
<td>Car use (%)</td>
<td>45.0</td>
<td>57.3</td>
<td>55.4</td>
</tr>
<tr>
<td>Public Transport use (%)</td>
<td>40.0</td>
<td>34.4</td>
<td>34.8</td>
</tr>
<tr>
<td>Walking use (%)</td>
<td>12.0</td>
<td>8.1</td>
<td>9.5</td>
</tr>
</tbody>
</table>

*Source: INE (2011) and Monzón et al. (2013)*

When examining the database it was found a significant share of people with a driving license having daily car availability, and a very low percentage of survey participants with a public transport monthly travel pass. This means that the sample is more private transport-oriented than the average population of the MMA. A simple statistical analysis (Table 4) reveals relevant changes after the implementation of the new transport policies. The only mode that lose market share was the car (from 57.3 to 55.4%) whilst walking increased the most (from 8.1 to 9.5%); this is mainly due to the economic crisis (Monzón et al., 2013). Moreover, travel time and commuting distance decreased between waves.
CHAPTER 3: STRUCTURAL EQUATION MODELS APPROACH
Structural Equation Modeling is an extremely flexible linear-in-parameters multivariate statistical modeling technique. It has been used in modeling travel behavior and values since about 1980, and its use is rapidly accelerating, partially due to the availability of improved software” (Golob, 2003)

After justifying the treatment of subjective variables on travel behavior in Madrid case study through a DCM (subchapter 2.1), it is necessary to catch the relationships between variables on travel behavior using SEM theory:

- First, with a complete study with socioeconomic, psychological, attitudinal and mobility variables using measurement (8) and structural equations (7) (subchapter 3.1).
- Then, with a factor analysis to catch built environmental factors on travel behavior (subchapter 3.2).

Subchapter 3.1 – Understanding travel behavior to change intentions towards private car and public transport

Traditional mobility studies measures travel behavior as: travel distance, travel time, travel mode choice, travel frequency, travel purpose, etc. It is often difficult to determine from the literature the order of importance or direction of causality of all these variables and it is highly likely that this will differ according to an individual’s personal circumstances and in different physical contexts. Moreover, it is difficult to find studies that deal with pro-car intention. The literature mainly concentrates on car use rather than car ownership, but it is not immediately evident that car use and ownership are driven by the same factors. Car ownership has been studied less frequently than car use and from fewer theoretical perspectives.

As shown by the discrete choice theory (Ben-Akiva and Lerman, 1985), frequency of use of one or other a transport mode can be statistically explained by variables such as place of residence, workplace, income, age, gender, etc. This theoretical perspective, much used by disciplines such as economics and transport engineering, grants special importance to elements like trip time or the trip cost (Salon, 2009; Vega and Reynolds-Feighan, 2009) and, in general, it assumes that people choose the transport mode that provides them with the highest utility or relative advantage (Domencich and McFadden, 1975; Dong et al., 2006).

Axhausen and Gärling (1992) observe that travel behavior models solely based on the Random Utility Theory (RUT) principle are not sufficient to understand how people make decisions. Although it has long been recognized that urban travel is a demand derived trough the need and desire for activity participation, RUT models do not incorporate it. Despite the growing interest in activity-based modeling within transportation research, a better understanding of the relation of causality for how urban form influences travel behavior is necessary. For example, a more distant shopping center may be chosen for its larger variety or better prices of products (Mokhtarian and Salomon, 2001). To put it in other words, there are travel behavior situations in which a more distant destination is chosen, because the utility of its inherent attractiveness exceeds the disutility of travel.

In this subchapter, the author focus on travel behavior into a comprehensive framework to capture the joint relationships between exogenous and endogenous variables to contrast the hypothesis that pro-car intention is the most relevant variable influencing negatively on pro-public transport intention instead of population’s household responsibilities or public transport level-of-service. The combination of the direct and indirect effects is known as the total effect of one variable on another in a Structural Equation Model
Assessment Methodology Applied to Demand Measures to Change Urban Mobility Behavior

(SEM). This study is one of the few applications of SEM in the understanding of travel behavior. In addition, to achieve a complete understanding travel behavior PhD candidate has broken down by modal choice the “travel behavior” considered in general terms for studies above. To validate the methodology, it uses data from a Madrid smartphone-based survey (n = 255 respondents) where some public transport policies have been introduced in the last two years: five new metro stations, by 25 % average public transport fare rise, on-street parking fee extended for one hour, etc.

Then a review of the theoretical and practical aspects to understand travel behavior has been made, i.e., a review of psychological and transport studies to capture the direct and indirect effects on travel behavior. Third section introduces the general assumptions of SEM, while the fourth section describes the panel smartphone data. A synthesis and the results applying the SEM methodology are presented in the fifth section, finalizing with some conclusions and recommendations.

Exogenous and endogenous variables on travel behavior studies

It is not easy to find in the literature a universal definition of exogenous and endogenous variables that define travel behavior. The major problem of similarity measurement is the lack of a generally accepted procedure to identify similarity of activity/travel patterns over long periods up to the meaningfulness of using of sequence alignment methods. Throughout following paragraphs the author reviews the most relevant variables used to capture the direct and indirect effects on travel behavior.

Exogenous variables

Scheiner and Holz-Rau (2007) used cross-sectional data collected in Cologne, Germany to investigate the relationships between life situation (socio-economic and demographic aspects), lifestyle (preferences and attitudes), choice of residential location (density of supply, quality of public transport and mixed land-use) and travel mode. The effect of location attitudes on travel behavior (travel mode choice and distance) is found to be equal or even stronger than the effects of residential location attributes on travel behavior. As a result, it contemplates the necessity to consider a combination of socio-economic and psychological dimension to explain travel behavior as exogenous variables.

- **Socio-economic variables**: Some of the arguments that explain the motivations that lead people to use the car rather than other means of transportation are based on the fact that adequate public transportation is not always available, and the car allows one to travel in less time and also to carry a piece of luggage, or comfort aspects for example (Jakobsson et al., 2002; Jakobsson, 2007; Cao et al., 2010). In contrast, symbolic motivations that intensify the use of the car refer to the fact that, through processes of social comparison, the car allows people to display their status and some aspects related to their social identity and self-concept (Steg, 2005; Gatersleben, 2007).

Some authors study interactions between household responsibilities variables (gender, age, number of children, etc.) and activity-travel behavior. Golob and McNally (1997) found that while out-of-home work activities limited male and female participation in other activities, complex relationships emerged with respect to gender-based activity trade-offs. Male recreational participation simultaneously reduced female discretionary activities while increasing female maintenance and travel activities. The presence of children seemed to encourage work and maintenance activity trading between male and female heads. Moreover, females were shown to have complex trip chains when compared with males. Perhaps the strongest link between travel behavior and gender was found by Polk (2003, 2004) in studies of travel behavior in Sweden in
1996. Women were more willing to reduce their use of the car than men, more positive towards reducing the environmental impact of travel modes and more positive towards ecological issues.

Household composition and income were also found to be major influences on travel behavior in a number of papers. Ryley (2006) studied the composition of 2910 households in Edinburgh. Ryley’s research showed that households with children have distinct travel behavior characteristics. These households are highly dependent on cars as the primary source of travel mode. Families consisting of retirees and high-income owners are least likely to use non-motorized forms of transport. Other approaches include these social status variables to improve the indirect effect of travel behavior. Kuppam and Pendyala (2001) indicate that income and ownership of a car or bicycle tended to favor out-of-home over in-home recreation.

Beyond the typical analysis of socio-demographic and activity-travel behavior, Golob (2000) set out to investigate the role of spatial accessibility. The results suggest that work travel times were also shown to improve with greater spatial access. Timmermans et al. (2003) signs that spatial context and urban form impact household activity participation and related travel costs.

However, the author likes to point out, that using socio-economic characteristics do not mean that there exist a direct causality between this variables and travel behavior. Nevertheless, using this kind of information as indicators or mediators of perceptions and attitudes behind the decision making process is common sense and often seems to be the only way to treat investigations based on revealed preference data.

- **Variables to describe attitudes and perceptions:** Some studies (Bagley and Mokhtarian, 2002; Schwanen and Mokhtarian, 2005a, b; Lois and López-Sáez, 2009; Van Acker et al., 2010) suggest that socio-psychological characteristics, such as attitudes, perceptions and preferences, may add explanatory power.

  Hiscock et al. (2002) studied the perceived psycho-social benefits of car use and ownership. In particular the authors studied the significance of the car as providing protection, autonomy and prestige compared with public transport. The results found that there were some psycho-social benefits to car users: their cars provided protection from ‘undesirable’ people, autonomy, convenience and greater access to a greater range of destinations than public transport.

  Ory and Mokhtarian (2004) analyze the factors that make people consider a trip to be more or less attractive. Along with objective aspects of the trip (such as the kilometers covered), these authors include attitudinal variables. Three variables in this latter category were the most powerful predictors of the variable attraction to travel behavior: (1) the practical benefits of the trip (traveling is not perceived as empty time, but as something useful and advantageous), (2) the sense of freedom (being able to travel wherever one wishes), and (3) status seeking through the trip (the opportunity that daily trips provide for individuals to display a major symbol of opulent and luxurious consumption).

**Endogenous variables**

- **Travel time.** Generally, travel time has been presented as negative effect on travel behavior. Transport studies are based on the assumption that demand for travel exists only because travelers want to reach their destination. Ory and Mokhtarian (2009) used more than 1300 commuters in San Francisco Bay Area to explain the connections among perceptions, affections and desires of short-
distance travel. They found that the shorter distances travelled by urban dwellers is relatively stable (and less influenced by enjoyment) regardless of the perceptions and affections.

**Number of trips.** It is a usual travel behavior indicator (Golob, 2000; Lee et al., 2007) but neither considers the temporal dimension of activity chains, nor the complexity of behavior. A lot of possible combinations of any two attributes (mode-purpose or purpose-destination, for example) are hardly performed (Outwater et al., 2003). It is necessary to establish at least the number of trips per mode and/or per purpose to take more into account the complexity of behavior and to complete this indicator by other characteristics. Schlich and Axhausen (2003) studied that the number of different trips characterized by more than two attributes does not increase much the performance of their models.

- **Number of home-based tours.** The relationship between out-of-home/in-home substitution effects and travel behavior was the subject of several studies. Lu and Pas (1999) measured travel behavior including a *trip chains* variable: a series of short trips linked together between anchor destinations. The rather intuitive finding was reported that increased in-home activity participation generated more trip chains and served to reduce time spent on out-of-home activities. McGuckin and Murakami (1999) established the difference between *trip chains* and *tours*: A tour is composed of “total travel between two anchor destinations, such as home and work, including both direct trips and chained trips with intervening stops”. Therefore, a possibility to study out-of-home/in-home substitution effects is to consider the number of home-based tours.

- **Day-to-day variability.** If travel is understood as a derived demand it is necessary to analyze variability in a complex way. People’s needs and desires vary from day-to-day. For example, the number of trips one takes varies from day-to-day because one does not need the grocery shopping each day. Moreover, behavior varies from day-to-day because of feedback from the transportation system. For example, in the event of congestion on the way to work, the day after a different route and/or departure time will be chosen. Several authors used GPS Data from multiday surveys to measure this variability with the definition of accurate and original variables (Li et al. 2004, Bath et al. 2005 and Mazloumi et al. 2009). On the other hand there exist habits that lead to similarity patterns during a week or month. Data based on multiday surveys allows studying this dimension of travel behavior. The different methods all report a similar trend concerning variability at different types of day: travel behavior is clearly more stable on working days.

**Car ownership: exogenous or endogenous variable on travel behavior?**

In some studies car ownership has been considered as an endogenous variable which is explained by a latent construct built by socio-economic variables. Dargay and Hanly (2004) indicate than highly educated people are found to own more cars and Kockelman (1997) establishes than car ownership is higher across high-income groups as well. On the other hand, car ownership can be considered as an exogenous variable together with socio-economic variables to explain travel behavior. As car ownership is related to income, car use will be higher across high-income groups (Dieleman et al., 2002; Rajamani et al., 2003).

Kockelman (1997) established as endogenous variables on travel behavior: dummy variables mode choice, distance and number of trips and car ownership across a sampled San Francisco Bay Area households. Golob and Meurs (1998) modeled travel time, vehicle miles of travel and car ownership together as endogenous variables on travel behavior, using data for Portland, Oregon. Simma and Axhausen (2003) demonstrate relationships between male and female heads of household with regard to
travel demands. The endogenous variables were car ownership, distances traveled by males and females, and male and female trips by two types of activities. Exogenous variables included the employment status of each head of household, family characteristics, and measures of residential accessibility and local land use.

Van Acker and Witlox (2010) confirmed the intermediary nature of car ownership as exogenous variable on travel behavior. Ben-Akiva and Atherton (1977) justified theoretically car ownership as mediating the relationship between the built environment and travel behavior: it is a medium-term decision influenced by other long-term decisions related to land use (place of employment and residential location choice). On the other hand, some studies consider car ownership directly as an exogenous variable and try to explain it based on various spatial and socio-economic variables (Bhat and Guo, 2007; Cao et al., 2007a; Giuliano and Dargay, 2006). Nowadays, only a limited amount of studies combines both research approaches and considers car ownership as mediating the relationship between the built environment, socio-economic variables and travel behavior (Cao et al., 2007b; Scheiner and Holz-Rau, 2007).

**Travel behavior as pro-car and/or pro-public transport intention**

Theories and research on attitudes, social norms and intention structure, and the relation between intention and behavior, are essential for an understanding of the choice to act pro-public transport (pro-environmentally). Different theoretical perspectives have been employed to describe the relation between intention factors and behavior. Two frequently applied frameworks in relation to travel behavior are: (i) a normative framework emphasizing altruistic moral motivation for traveling pro-environmentally (Norm Activation Model, NAM) (Schwartz, 1977); and (ii) a rational choice framework stressing self-interest and utilitarian aspects when explaining travel behavior (Theory of Planned Behavior TPB) (Ajzen, 1991). NAM theory shows an activated personal norm to reduce the negative environmental effects of car use is a morally based motivation to act for the sake of the environment when making decisions about how to travel. TPB stipulates that an intention to travel pro-environmentally, for example using public transport, is a result of various considerations expressed mainly in: an attitude, subjective norm, rather than a perceived internal pressure to save the environment as stipulated in the NAM.

Several researchers advocate a combination of the TPB and a normative model (NAM) in order to explain travel behavior (Harland et al., 1999; Wall et al., 2007). This conclusion is supported by studies within the social dilemma framework, where travel behavior has been found to be motivated by a combination of self-interest (pro-car intention) and concern for others (pro-public transport intention) (Van Lange et al., 1998).

**Formulation of SEM and mediation effect**

SEM enables the estimation of bi-directional relationship (feedback loops) between variables whereas regression allows only unidirectional relationship. Thus, the SEM is superior to linear regression. Nevertheless, SEM based on similar techniques so that the researcher has to pay attention to the same assumptions like linear regression.

Figure 8 shows the structure of SEM in a general way. Next to a measurement model of exogenous latent variables ($\xi$) exists a similar model for endogenous latent variables ($\eta$). Both measurement models are connected by assumed causality relationship.
Exogenous parts of SEM could be defined as reflective or formative constructs. In reflective constructs the latent construct is the cause of the measured indicators. Indicators are interchangeable among each other. That means that a change of the latent construct is shown through a change (reflection) in all indicators. On the other hand, a formative construct is understood as a weighted composition of indicators. The specific behind this formative construct is to understand the indicators as a cause of a latent construct. In this case, the indicators are not necessarily highly correlated among each other like in a reflective measurement model. For this reason, a change in one of the indicators leads to another latent construct. The decision on which type is appropriate is exclusively a question of content. The literature gives some decision rules to establish a correct type of model (Christophersen and Grape, 2006).

Following the four-step approach (Mulaik and Millsap, 2000) presented at chapter 1.3 to testing a nested sequence of SEM.

**Mediation Effect**

Above deals with mediation effect of car ownership variable between exogenous variables and travel behavior. The PhD candidate considers that this methodology approach is necessary for this study. There are three main types of mediation effect (Robins, 1992): i) indirect, ii) partial and iii) full. Indirect means that the direct effect never was significant, but that the indirect effect is. Partial mediation means that both direct and indirect effects from the X (exogenous) to Y are significant. Full means that the direct effect drops out of significance when the mediator (M) is added, and that the indirect effect is significant. The figure 9 illustrates these types of mediation.
Definition of variables and results

The data use in this study originates from first wave (n=255 respondents) of the HABIT project (Habit and Inertia in mode choice behavior: a data panel for Madrid).

Based on the findings of previous studies presented in chapter 2 above and the data available, seven socio-demographic variables are included as exogenous variables of travel behavior:

- SEX (Male = ‘1’; Female = ‘0’)
- AGE
- INCOME (monthly household income)
- CHILD (presence of child = ‘1’, no = ‘0’)
- FAMILY (number of family members)
- HELPED (if he/she received some help for child-care or for housekeeping = ‘1’, no = ‘0’)
- EMPLOYMENT (full-time employed = ‘1’; otherwise = ‘0’)

At each trip, perceptions towards public transport are measured using 4-point Likert scale:

- ACCESS_TIME
- PHYSICAL_ACCESIBILITY
- TRAVEL_TIME
- WAITING_TIME
- TRANSFER
- SPACE
- SAFETY
- CLEANLINESS

Due to the fact that the perception towards public transport is asked for each trip the average perception as arithmetic mean is used assuming that the Likert scale can be treated as a continuous variable (Wang and Law, 2007).

As for endogenous variables, it is able to calculate with data availability the most usual travel behavior indicators –or LOS indicators- presented above:

- TIME: weekly travel time in minutes.
- NUMBER_TRIPS: weekly number of trips.
- TOURS: number of weekly home-based tours. To consider TOURS by mode, it has taken into account the most representative choice.
Assessment Methodology Applied to Demand Measures to Change Urban Mobility Behavior

- SP: day-to-day variability. There are a lot of approaches for considering day-to-day variability (see review of Schlich and Axhausen, 2003; and Buliung et al., 2008). These authors consider among others the index SP (Similarity index Pas) which compares the trips of a day pairwise (Pas, 1983). This laborious index has had to be implemented with the MATLAB language because trips are compared in their order of occurrence and it is hard to do this by hand.

By using data from a five-days travel diary in a SEM, temporal fluctuations weekly can be reduced while allowing calculation variability for different types of day with TOURS and SP variables. Finally, CAR_OWN variable (car ownership = ‘1’, no = ‘0’) is considered as a particular mediation effect mentioned between exogenous and endogenous variables.

The SEM methodology –presented at subchapter 1.4- is used to investigate links between both objective and subjective personal variables and travel behavior, and follows the four step estimation approach recommended by Mulaik and Millsap (2000). The statistical program SPSS is used in step 1 to perform EFA through the default method of extraction in other statistical applications: Principal Component Analysis (PCA). Moreover, the AMOS module of the SPSS software package (Arbuckle and Wothke, 1999) is used in steps 2, 3 and 4.

- **Step 1. EFA, Measurements Models:** The first step looks at the correlation matrix appearing a very low (<0.3) correlation factor between SEX and the rest of the variables, as well as EMPLOYMENT; thus, it can reject these variables (Cohen, 1992).

To run a first PCA, variables must fulfill the following assumptions: reliability (Cronbach’s alpha), homoscedastic, positive correlations between variables and normality, (Field, 2009). The Cronbach’s alpha for the 21 resulting variables is -0.030, a very low internal consistency. Therefore, it is useful a transformation of the scale variables. For example, logarithmic transformation is used to reach the assumptions above. Moreover, it identified a negative correlation between LOS variables corresponding to the two different modes: CAR and PT. As first approach, it left out of account this aspect, but in step 3 this limitation will be solved. Now the Cronbach’s alpha is 0.808, a very acceptable value. Running this first PCA, all the variables did not reach normality assumption (dummy variables, for example), but a long list of travel behavior studies with SEM approach exists, using ordinal and dummy variables with satisfactory results (see review of Golob, 2003). Finally this first PCA analysis, the scree-plot indicates a clear “elbow” with 4 factors (scores) meaning that a 3-factor-solution can be extracted (Field, 2009). The last factor (without rotation) has an eigenvalue of 1.813 and explains 69.6 % of the variance.

A second run of PCA with a fixed factor extraction of 3 leads to very small communality SP (0.046). The personal variables which built the latent factor “household responsibility” shows communalities, except CHILD less than 0.5 (0.302 – 0.490); but they are useful to design a measurement model for this latent factor since the factor loadings have values higher than 0.5 which means that they are necessary to perform the latent factor “household responsibility” (Field, 2009).

A third run of a PCA without SP variable leads to an acceptable result. This three-factor-solution now describes 72.9 % of variance of the used variables. The PCA reach an excellent valuation of the Kaiser-Meyer-Olkin-Criteria (KMO=0.907). The Bartlett-Test indicates a high significance level (\(\chi^2\) (190) = 5.052, \(p < 0.0001\)) and the Anti-Image-Matrix is near to zero for all of the values. The three latent factors could be named as “household responsibility” (HH_RESP), “comfort
perception” (COMFORT) and “travel behavior” (TRAVEL_BEHAVIOR). Therefore, three measurement models are extracted by PCA to check by the next step.

- **Step 2. CFA. Modification Indices:** Once the number of latent constructs (HHRESP, COMFORT, TRAVEL_BEHAVIOR) has been fixed in each measurement model, it consider model modifications to achieve a better model fit values. If a measurement model has model fit indices that are less than satisfactory (a RMSEA less than 0.5), a researcher typically performs a specification search to find a better fitting model to the sample variance-covariance matrix (Schumacker and Lomax, 2004). To include additional parameters –to improve the goodness of fit for each model- the most commonly used technique is select the highest modification index (MI) (the expected value that $X^2$ would decrease by if such a parameter are to be included). Measurement model results are showed in Figure 9 below.

**Figure 9. Measurement models including factor loadings, correlations and model fit values**

**Measurement models of exogenous latent variables:** HH_RESP and COMFORT

- RMSEA=0.012; CFI=0.999; AGFI=0.899

**Measurement model of endogenous latent variable: TRAVEL_BEHAVIOR**

- RMSEA=0.049; CFI=0.998; AGFI=0.984

- **Step 3. Structural Model:**
  
  **Model Specification.** It hypothesizes a structural equation model based on predicting pro-public transport intention (PROPT) as a latent dependent variable to explain travel behavior. The structural model is diagrammed in Figure 10 with four latent variables (ellipses): two latent independent variables, household responsibilities (HH_RESP) and comfort perception (COMFORT), two latent dependent variables, pro-car intention (PROCAR) and pro-public transport intention (PROPT), and car ownership (CAR_OWN) as mediation variable between HH_RESP and PROCAR and PROPT latent variables. Therefore, the most important hypotheses that modify the results of measurement model included by the PhD candidate are two: (1) to consider CAR_OWN as mediation variable and (2) to separate TRAVEL_BEHAVIOR latent variable into PROCAR and PROPT dependent latent variables.
In Figure 10 above each observed variable (rectangles) has a factor loading (i.e., that part of each observed variable is measuring something other than the hypothesized latent variable) and a unique measurement error that forms an equation to compute the latent variable score; for example, $HH\_RESP = f(CHILD, INCOME, AGE, HELPED, FAMILY) +$ measurement error. Moreover, each latent dependent variable has one or more structure coefficients and a unique prediction error (i.e., that part of latent dependent variable is not predicted by the independent variables) that goes into an equation; for example, $PROCAR = f(HH\_RESP + CAR\_OWN + COMFORT) +$ prediction error. There are two equations in our hypothesized structural model, so two prediction errors are estimated, one for $PROCAR$ and one for $PROPT$.

Model Identification. It determines the model identification by first checking the order condition: the number of values in sample variance-covariance matrix has to be greater than the number of free parameters.

Model Estimation. As stated above, the technique of Maximum Likelihood (ML) to estimate the SEM is used. To validate all the relationships hypothesized in Figure 10, it is necessary to study the mediation effect (indirect, partial or full mediation) with an appropriate technique. Bootstrapping is a resampling method that creates a sampling distribution to estimate standard errors and create the confidence intervals. Apart from being an aid to nonnormal data, it is important to confirm the mediation effect of a variable: accuracy for computing confidence intervals for mediation effect when the mediation effect is nonzero (Cheung and Lau, 2008). Results above indicate that five mediation variables based on the PhD candidate’s hypothesis exist:

- Partial Mediation I: $HHRES \rightarrow PROCAR \rightarrow PROPT$
- Partial Mediation II: $COMFORT \rightarrow PROCAR \rightarrow PROPT$
- Partial Mediation III: $HHRES \rightarrow OWN \rightarrow PROCAR$
- Partial Mediation IV: $COMFORT \rightarrow OWN \rightarrow PROCAR$
- Partial Mediation V: $OWN \rightarrow PROCAR \rightarrow PROPT$

Table 5 summarizes the results of a mediation study with bootstrapping based on evaluating the significance of standardized regression weights between: $HH\_RES \rightarrow PROPT$ (Mediation I), $CAR\_OWN \rightarrow PROPT$ (Mediation II), $HH\_RES \rightarrow PROPCAR$ (Mediation III),...
COMFORT→PROPCAR (Mediation IV) and CAR_OWN→PROPT (Mediation V); without and with a Med variable (OWN or PROCAR) to evaluate the type of mediation observed.

Table 5. Mediation study (Bootstrapping method)

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Direct effect without Med (Total effect)</th>
<th>Direct effect with Med</th>
<th>Indirect effect</th>
<th>Mediation observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partial Mediation I</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HHRES→PROCAR→PROPT</td>
<td>-0.238 **</td>
<td>0.042</td>
<td>-0.284*</td>
<td>Full</td>
</tr>
<tr>
<td>Partial Mediation II</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COMFORT→PROCAR→PROPT</td>
<td>-0.553 **</td>
<td>-0.302 **</td>
<td>-0.281**</td>
<td>Partial</td>
</tr>
<tr>
<td>Partial Mediation III</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HHRES→OWN→PROCAR</td>
<td>0.388 **</td>
<td>0.217 **</td>
<td>0.172 **</td>
<td>Partial</td>
</tr>
<tr>
<td>Partial Mediation IV</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COMFORT→OWN→PROCAR</td>
<td>0.392 **</td>
<td>0.306 **</td>
<td>0.086 **</td>
<td>Partial</td>
</tr>
<tr>
<td>Partial Mediation V</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OWN→PROCAR→PROPT</td>
<td>-0.299 **</td>
<td>-0.081</td>
<td>-0.218*</td>
<td>Full</td>
</tr>
</tbody>
</table>

`.` p < 0.1; `*` p < 0.05; `**` p < 0.01 (two-tailed)

These results modified the relationships hypothesized in Figure 4. Thus, it does not exist a direct effect between HH_RESP and PROPT with PROCAR as mediation—it has to delete the two corresponding arrows—, supporting the great influence of PROCAR variable in the model.

Figure 11 shows ML estimation final results for the structural equations (structure coefficients) and measurement equations (factor loadings). All of the parameter estimates are significantly different from zero (p < .05).

Figure 11. Maximum Likelihood estimates for SEM

Pro-car intention is mainly influenced by car ownership (0.42). Pro-car intention is less complex among household responsibilities (0.22), and comfort perception (0.31). Higher household responsibilities (high income, greater number of children, older…) are associated with more trips, longer travel times, and longer number of home-based tours because logarithm function is a positively increasing function. Pro-public transport intention is negatively influenced directly or indirectly for all the exogenous variables of this study, but the most important aspect is that the lower pro-car intention (-0.68), the higher pro-public transport intention. However, examining direct effects only can be misleading. Analyzing each exogenous latent variable, CHILD
(presence of child=1) indicator has a major impact on the HH_RESP with a factor loading value of 0.92. The rest of indicators with a direct influence on HH_RESP present a factor loading values between 0.23 and 0.49. Similarly, WAITING_TIME with a factor loading of 0.98 represent the mostly explains the COMFORT variable.

- **Step 4. Model Testing:** When the model fit indices are acceptable, the hypothesized structural model has been supported by the sample variance-covariance data. For the model presented the \( \chi^2/df \), a measure of goodness of fit, is equal to 1.437 which indicates that the model is acceptable. The root-mean-square error of approximation (RMSEA) is equal to 0.046, also below the typical acceptable level of model fit (RMSEA<0.05). The comparative fit index (CFI) is 0.987, larger our acceptable range of model fit (CFI>0.95). Finally, the adjusted goodness-of-fit index (AGFI) is 0.879, around the acceptable range of model fit (AGFI>0.9). On the basis of these criteria, the model fits the data properly and it can conclude that the model meets our expectations regarding statistical adequacy. Therefore, the two hypotheses introduced by the author at the beginning of step 3 are correct statistically.

**Brief of principal travel behavior relationships**

The analysis is considered a starting point for the thesis. Therefore, more research is needed on the interrelationships between previous specific variables and others like attitudes (conveniences) and spatial variables to improve the exploration of the complexity of travel behavior. Further chapters will be oriented toward estimation of a hybrid demand model by using land use variables and other psychological factors distinguishing whether these are about attitudinal, affective, or social aspects.

Causal effects of socioeconomic and comfort perception variables on pro-public transport intention are much lesser than those negative causal effects of pro-car intention on pro-public transport intention. Overall, these results suggest that to encourage pro-public transport intention it is necessary to reduce pro-car intention reducing demand for car (time, number of trips and number of home-based tours, all for car use). That is, if intention of using private car would decrease, intention of using public transport would increase and (to a lesser extent because pro-car intention depends directly on other personal variables) vice versa.

Additionally, it can be seen from the bootstrapping method the great importance of the mediation treatment of “car ownership” in SEM to examine the causal relationships in travel behavior, because the relationship between “household responsibilities” and PROCAR has a total effect –including “car ownership” variable- of 0.388 but an indirect effect of 0.172. In the same way, the total effect between “comfort perception” and PROCAR has a total effect of 0.392 and an indirect effect of 0.086.

To sum up the results of the subchapter focus on transport policies:

- Transport policy actions are more likely to be effective when pro-car intention has been disrupted first because: (1) cars provide places where people can feel a positive “comfort” in contrast to public transport; and (2) the more “household responsibilities”, the more pro-car intention.

- Information campaigns and new or improved public transport infrastructures are unlikely to be sufficient to move people from considering an alternative mode of transport that increases the pro-public transport intention because “household responsibilities” changes directly on pro-car intention and indirectly on pro-public transport intention.
Subchapter 3.2 – Urban built environment analysis

Applications involving travel behavior from the perspective of land use are dating from the 1990s. Usually, four important components are distinguished: density, diversity and design (3D’s of Cervero and Kockelman) and accessibility (introduced by Geurs and van Wee). But there is not a general agreement on how to measure each of those 4 components. Density is used to be measured as population and employment densities, but others authors separate population density between residential and building densities. A lot of measures have been developed to estimate diversity: among others, a dissimilarity index to indicate the degree to which different land uses lie within one another’s surrounding, an entropy index to quantify the degree of balance across various land use types or proximities to commercial-retail uses. Design has been characterized by site design, and dwelling and street characteristics. Lastly, accessibility has become a frequently used concept, but its meaning on travel behavior field always refers to the ability “to reach activities or locations by means of a travel mode”, measured as accessibility to jobs, to leisure activities, and others. Furthermore, the previous evidence is mainly based on US data or on north European countries. Therefore, this subchapter adds some new evidence from a Spanish perspective to the research debate. Through the Madrid smartphone-based survey, factor analysis is used to linearly combine variables into the 3D’s and accessibility dimensions of the built environment. At a first step for the rest of the thesis, land use variables will be treated to define accurately the previous 4 components.

The relevance of built environment

Since 1970s the most frequently quoted studies on the impact of the land use patterns on travel behavior have provided important conclusions. Through a regression analysis, Hurst (1970) demonstrated that higher rates of vehicle trip generation were found among retail and office land uses compared with storage and industrial usage. Newman and Kenworthy (1989) found a significant negative statistical correlation between residential density and transportation-related energy consumption per capita.

The impact of density, diversity, accessibility, and percentage of multifamily residential on travel time was study by Ewing et al. (1994). Also Friedman et al. (1994) distinguished two neighborhood types: standard suburban and neo-traditional neighborhoods. To find a study that analyzes different neighborhood design, Hess et al. (1999) demonstrated that urban neighborhoods with small blocks and extensive sidewalk systems were found to generate three times more the pedestrian volumes than suburban sites with large blocks and short. Handy (1996) was among the first to mention the importance of perceptions and attitudes towards land use on travel behavior. But also there are studies that point to a higher significance of land use compare to socio-economic and demographic characteristics (Schwanen and Mokhtarani 2003, 2005).

How to measure spatial dimension?

Most evidence about how to measure the environmental dimension on travel behavior is based on data stemming from the US as stated before. Since 2000s, the research debate has been enriched with European evidence. But there are important differences in urbanization patterns between North-American and European cities. Thus, the great problem is the lack of a common point in all of the studies in this field to quantify this novelty dimension. In this part, the author review the ways used in different contexts to achieve the target of this subchapter: a quantification of the spatial dimension in a travel behavior study since Spanish perspective.
Key dimensions to be taken into account

Land use characteristics can be measured at several scales, ranging from the local neighborhood to the metropolitan area. Usually, four important components are distinguished: density, diversity and design (Cervero and Kockelman, 1997) and accessibility (Geurs and van Wee, 2004).

Density. The effects of density on travel demand have long been acknowledged (e.g., Levinson and Wynn, 1963) and remain well-studied and understood. Higher densities are associated with more public transport use, more walking and cycling, and less car use. After all, public transport is organized more efficiently (more routes, higher frequency of services) in high density areas and car users may face more congestion. Furthermore, travel distance and time is negatively associated with density (Cervero and Kockelman, 1997; Kitamura et al., 1997; Schwanen et al., 2004).

Diversity. Several measures have been developed to estimate diversity: among others, a jobs/housing ratio (Ewing et al., 1994), an entropy index to quantify the degree of balance across various land use types (Kockelman, 1997) or a dissimilarity index to indicate the degree to which different land uses lie within one another’s surrounding (Kockelman, 1997). The effects of more diversity on travel behavior are comparable to the effects of higher densities.

Design. The factor design can be characterized by a general classification of neighborhoods with a standard suburban neighborhood and a neo-traditional neighborhood as extremes (Gorham, 2002). Standard suburban neighborhoods are characterized by low densities, limited diversity, and a car-orientated design. However, design can be characterized more specifically by site design, and dwelling and street characteristics. Studies indicate that neighborhoods characterized by small block sizes, a complete sidewalk system, the absence of cul-de-sacs and limited residential parking encourage walking and cycling (Cervero and Kockelman, 1997; Hess et al., 1999).

Accessibility. Accessibility is a fourth important land use characteristic. Accessibility has become a frequently used concept, but its meaning always refers to the ability “to reach activities or locations by means of a (combination of) travel mode(s)” (Geurs and van Wee, 2004). To measure this accessibility there are two approaches: (i) according to Koenig (1980) it must take into account the distance between the person or place and the destination and the utility of various destinations; but also (ii) according to Simma and Axhausen (2001) accessibility is calculated as the number of reachable facilities. The latter approach is labeled as intensity by Krygsman and Dijst (2001). Moreover, most studies agree on the effects of accessibility on travel behavior. For example, Gao et al. (2008) found that households living in residential locations with higher job accessibility are likely to own fewer cars. Several studies also point out that accessibility is negatively associated with travel times (e.g., Ewing et al., 1994; Susilo and Maat, 2007).

Land use variables tested

The many ways by which urban environment can be measured may be considered as observed characteristics of a neighborhood. Several of these observed characteristics are related. In some cases the relationships are obvious: the average proximity to a transit stop in a neighborhood and transit-based accessibility to opportunities may be correlated in measuring overall transit access of a location. However, as the number of observed variables increases, it is difficult to identify the structure within them. Therefore it becomes necessary to condense these observed variables into a smaller set of variables that accounts for the variance in the data. One such data reduction technique is FA. Factors derived through such data reduction techniques are also referred to as latent variables. Although, as the name latent implies, these variables are not observable, certain effects on measurable (manifest) variables can be observed
Assessment Methodology Applied to Demand Measures to Change Urban Mobility Behavior

(Srinivasan, 2001). Thus, FA methods can assess and explain the structure in a set of correlated, observed variables in terms of a small number of latent variables or factors.

Cervero and Kockelman (1997) introduced the idea that “in light of the need to use sets of variables to capture the many-sided dimensions of built environments and to allow for collinearity, the multivariate technique of factor analysis was used”. These authors pointed that since the research focused on how the land use shaped travel demand, FA was carried out only for land use variables (not socioeconomic, attitudinal, etc.). Their FA is successful in providing a multi-variable description of two of the 3D’s dimensions (density and design) specified each other by six land use variables. Each factor is labeled ‘intensity’ and ‘walking quality’, respectively. It is the beginning of several studies in this field that included FA to improve their studies since the multicolinearity among the land use variables could hide the effects of their individual contributions to travel demand.

Since the great goal of this paper is to define new evidence from the Spanish perspective to the land use variables treatment on travel behavior, it is necessary to review how the land use factors have been labeled in other American and European studies following the technique of FA (Table 6).

<table>
<thead>
<tr>
<th>Study</th>
<th>Environmental Factors labeled</th>
</tr>
</thead>
<tbody>
<tr>
<td>de Abreu e Silva, Golob, and Goulas (2006)</td>
<td>(1) Residence in traditional urban areas; (2) Working in traditional urban areas; (3) Working in compact and central urban areas; (4) Road supply; (5) Freeway supply in the residence area; (6) Residence in a specialized area</td>
</tr>
<tr>
<td>Van Acke and Witlox (2010)</td>
<td>(1) Built up index; (2) Land use diversity; (3) Distance to railway station; (4) Distance to CBD; (5) Accessibility by car</td>
</tr>
<tr>
<td>de Abreu e Silva and Goulas (2009)</td>
<td>(1) Employment in a central and denser area; (2) Residence in a central and denser area</td>
</tr>
<tr>
<td>Ewing and Cervero (2010)</td>
<td>(1) Density; (2) Diversity; (3) Design; (4) Accessibility; (5) Distance to CBD</td>
</tr>
<tr>
<td>de Abreu e Silva, Goulas and Dalal (2012)</td>
<td>(1) Employment in a central, denser and accessible area; (2) Residence in a compact, denser and accessible area; (3) Employment in a dense area well served with roads; (4) Residence in a compact and small area and well served by roads; (5) Working in a mixed and compact zone; (6) Residence in a mixed and well served by freeways area; (7) Mix of land uses in the residence area</td>
</tr>
<tr>
<td>He and Zhang (2014)</td>
<td>(1) Density; (2) Entropy; (3) Average block size; (4) Distance to CBD</td>
</tr>
</tbody>
</table>

The six studies above confirm the lack of common point to define built environmental factors (latent variables) that are represented by a group of land use observed variables on transport studies. Among the general conclusions, noteworthy: (i) a density factor defined at least from population density; (ii) a diversity factor defined at least from land use mix value or entropy index; (iii) a design factor but defined on different ways (street quality factor or average block size, for example); and (iv) some accessibility factors defined since different points of view (accessibility by car, distance to transit or public transport, distance to CBD, residence/employment in a specialized area, etc.). Moreover, the variables used to define the factors are scale variables or/and percentage, being greater the number of the first ones. Most part of the previous studies finished with a Structural Equation Modeling approach (SEM) to analyze the relationship between those new land use factors and travel behavior. This type of methodology will be applied in this subchapter.

Formulation of Factor Analysis

The utility of FA hinges on its ability to yield stable, accurate and interpretable estimates of factor loadings. The mathematical formulation was evaluated at subchapter 1.4 as measurement equations (8), but there are a number of determinants of successful application of FA that have to be taken into account following the next stepped way proposed by the PhD candidate:
**Step 1. Data suitable.** Although sample size is important in FA, there are varying opinions, and several guiding rules of thumb are cited in the literature. Hair et al. (2009) suggested that sample sizes should be 100 or greater. A number of textbooks cite the work of Comrey and Lee (2013) in their guide to sample sizes: 100 as poor, 200 as fair, 300 as good, 500 as very good, and 1000 or more as excellent. Moreover, Tabachnick and Fidell (2007) recommended inspecting the correlation matrix for correlation coefficients over 0.30. Prior to the extraction of the factors, several tests should be used to assess the suitability of the respondent data for FA, as Kaiser-Meyer-Olkin (KMO) index. The KMO index ranges from 0 to 1, with 0.50 is considered suitable for FA.

**Step 2. Factor extraction method.** The aim of the extraction is to simplify the factor structure of a group of items, or in other words, high item loadings on one factor and smaller item loadings on the remaining factor solutions. There are numerous ways to extract factors: Principal Component Analysis (PCA), Principal Axis Factoring (PAF) and others. PCA and PAF are used most commonly in the published literature (Thompson, 2004; Tabachnick and Fidell, 2007). When the variables have high reliability (Cronbach’s alpha) the differences between the two are often insignificant (Thompson, 2004); but PCA is recommended when no priori theory or model exists (Gorsuch, 1983).

**Step 3. Rotational Method.** Another consideration when deciding how many factors are analyzed the data is whether a variable might relate to more than one factor. Rotation maximizes high item loadings and minimizes low item loadings, therefore producing a more interpretable and simplified solution. There are two common rotation techniques: orthogonal rotation and oblique rotation. Orthogonal rotation first developed by Thompson (2004) is the most common rotational technique used in FA, which produce factor structures that are uncorrelated. In contrast, oblique rotation produce factors that are correlated, which is often seen as producing more accurate results for research involving human behaviors, or when data does not meet priori assumptions (Costello and Osborne, 2005).

**Step 4. Number of factors.** The aim of the data extraction is to reduce a large number of items into factors. In order to produce scale unidimensionality, and simplify the factor solutions several criteria are available to researchers. However, given the choice and sometimes confusing nature of FA, no single criteria should be assumed to determine factor extraction (Costello and Osborne, 2005). According to Hair et. al (2009) factors should be stopped when at least 50-60% of the variance is explained (for social sciences).

**Step 5. Interpretation.** Interpretation involves the researcher examining which variables are attributable to a factor, and giving that factor a name or theme. Traditionally, at least two or three variables must load on a factor so it can be given a meaningful interpretation (Henson and Roberts, 2006). Variables with higher loadings are considered more important and have greater influence on the name or label selected to represent a factor. The signs are interpreted just as with any other correlation coefficients. If the researcher is content with these factors, these should be operationalized and descriptively labeled. It is important that these labels or constructs reflect the theoretical and conceptual intent.

**Definition of variables and results**

The variables related to the service area have been calculated within the station service area (SSAs). Service areas were obtained using Geographical Information System (GIS), and are based on distances across the transport/road network. The distance threshold considered was 900 meters, which is the maximum distance that most people are willing to walk in order to access the Metro network in Madrid (García-Palomares et al., 2013). Once the SSAs were defined with a GIS, they have been intersected with
Assessment Methodology Applied to Demand Measures to Change Urban Mobility Behavior

various urban variables that that hypothetically favor transit use: density, diversity and design (Cervero, R. and Kockelman, 1997); and accessibility dimensions (Geurs and van Wee, 2004).

Population Density and Employment Density have been chosen as density variables in the service area. The former was calculated as inhabitants/ha, and the latter as employment/ha. Two indicators of “land use mix” have been used. First, the ratio of employment per inhabitant was computed (Job Ratio). This index can contribute to measure the job accessibility. Second, a more general land use mix (Mix) was measured using the reciprocal of the variation coefficient of the area covered by different land uses within SSAs (higher values indicate higher diversity in uses). Both measures are easily computable and interpretable. But the rest of the different categories of land uses within SSAs have been also studied separately to improve the knowledge about the diversity and accessibility dimension: hectares (ha) of trade, health and educational (Equipment), ha of single-family residential (Single Residential), ha of multifamily residential (Multi Residential), ha of industry (Industry), ha of Offices, ha of infrastructure that promotes economic activity, such as roads, highways, railroads, airports, electricity, telecommunications, water supply and sanitation (Infrastructures), ha of parks and recreations (Green Zones). To measure the “center accessibility”, the distance of each SSAs to Center Business District (Distance CBD) was also included, as well as ha of land available for building (Brownfield).

An urban design indicator was calculated using the street network layer Street Density within SSAs. This variable was calculated as a ratio between the street length and the service catchment area. Street Density can be considered as an indicator of walkability (Zhu and Lee, 2008), since it favors access to stations on foot and increases transit ridership (Cervero, 2002). To achieve a better knowledge of “street network design” dimension was used the inputs of Ravulaparthy and Goulias (2014) that set of centrality measures to spatial systems:

- Remoteness centrality: measures to what extent a link is close to all the other links along the shortest paths from one link to another on the network.
- Betweeness centrality: is based on the idea that a link is more central when it is traversed by a large number of shortest paths connecting any other two links in the network.
- Straightness centrality: represents “efficiency of communication” between two links increases when there is a least deviation of their shortest path from the virtual straight line connecting them – that is, a greater straightness of the shortest-path distance.
- Reach centrality: measures the number of other links that can be reached along the shortest path on a network.

This centrality measures complement the classical Multiple Centrality Assessment (MCA) model (Porta et al., 2012) in two ways: (a) accommodate the context of location and its importance through weighted link attributes like roadway capacity, population and opportunities at a place; and (b) accounting for the relative importance of a link in the network across multiple spatial scales and centrality values. To determine at least the small-scale measures for the centrality indices above, it was computed centrality indices for a network radii or network buffer surrounding each link of 2.5km along with measures for the entire Madrid network, which are the 25th percentile, of the pairwise distance distribution.
Table 7. FA results

<table>
<thead>
<tr>
<th>Factor</th>
<th>Observed variable</th>
<th>Factor loading</th>
</tr>
</thead>
<tbody>
<tr>
<td>Street Network Design</td>
<td>Straightness</td>
<td>0.970</td>
</tr>
<tr>
<td></td>
<td>Reach</td>
<td>0.961</td>
</tr>
<tr>
<td></td>
<td>Remoteness</td>
<td>0.947</td>
</tr>
<tr>
<td></td>
<td>Betweenness</td>
<td>0.805</td>
</tr>
<tr>
<td>Urban Block Diversity</td>
<td>Mix</td>
<td>0.798</td>
</tr>
<tr>
<td></td>
<td>Multi Residential</td>
<td>0.773</td>
</tr>
<tr>
<td></td>
<td>InvLog_Road Supply</td>
<td>0.605</td>
</tr>
<tr>
<td>Nonresidential Diversity</td>
<td>Green Zones</td>
<td>0.772</td>
</tr>
<tr>
<td></td>
<td>Industry</td>
<td>0.725</td>
</tr>
<tr>
<td></td>
<td>Infrastructures</td>
<td>0.723</td>
</tr>
<tr>
<td>Job Accessibility</td>
<td>Employment Density</td>
<td>0.898</td>
</tr>
<tr>
<td></td>
<td>Job Ratio</td>
<td>0.675</td>
</tr>
<tr>
<td></td>
<td>Offices</td>
<td>0.561</td>
</tr>
<tr>
<td>Center Accessibility</td>
<td>Brownfield</td>
<td>0.832</td>
</tr>
<tr>
<td></td>
<td>Distance CBD</td>
<td>0.734</td>
</tr>
<tr>
<td>Density</td>
<td>Population Density</td>
<td>0.691</td>
</tr>
<tr>
<td></td>
<td>InvLog_Single Residential</td>
<td>0.666</td>
</tr>
</tbody>
</table>

Trying to identify a combination of variables into the 3D's (density, diversity and design) and accessibility dimensions of the built environment, results above add some new evidence from a Spanish perspective to the research debate. Each factor loading represents the role each observed variable plays in defining each factor:

- as more *Straightness, Reach, Remoteness* and *Betweenness* values, bigger ‘street network design’.
- as more *Mix, Multi Residential* values, bigger ‘urban block diversity’. Conversely, as more ‘road supply’ value, smaller this diversity factor.
- as more *Green Zones, Industry, Infrastructure* values, bigger ‘nonresidential diversity’.
- as more *Employment Density, Job Ratio, Offices* values, bigger ‘job accessibility’.
- as more *Brownfield* and *Distance CBD* values, bigger ‘center accessibility’, i.e., more difficult to access to the city center.
- as more *Population Density* value, bigger the ‘density’ factor. Conversely, as more Single Residential value, lower density factor.

The goal of a FA is to identify a limited number of underlying (latent) factors responsible for observed variances and covariances. Following the stepped methodology proposed at chapter 4 and the database presented at chapter 5, it tries to identify dimensions of the urban built environment since Spanish perspective.

**Step 1. Data suitable.** With an acceptable sample size of 255 respondents, the correlation matrix of the eighteen environmental variables presents a very low (<0.3) correlation factor between Equipment and the rest of the variables; thus it can reject this variable. To assess the suitability of the respondent data for the FA with the seventeen resulting variables it reaches an acceptable valuation of KMO index (0.684).

**Step 2. Factor extraction method.** The statistical program SPSS is used to perform FA through the default method of extraction in other statistical applications: Principal Component Analysis (PCA). Moreover PCA is appropriate when the primary concern is about prediction or the minimum number of factors needed to account for the maximum portion of the variance represented in the original set of variables, and when no priori theory exists.
Step 3. Rotational Method. Reviewing similar environmental-factor-analysis researches based on different case studies (see chapter 3) the factors will be related; thus the oblique rotation method is more suitable.

Step 4. Number of factors. With a first PCA using oblique rotation method, the scree-plot indicates a clear “elbow” with seven factors (scores) meaning that a six-factor-solution can be extracted. Communalities are uniformly high (between 0.6 and 0.9), but two observed variables (Road Supply and Single Residential) have negative factor loadings. Thus, it is useful a transformation of the scale variables. To achieve factor loadings more than 0.5 it is necessary to apply the inverse of the logarithm for both variables. Running a second PCA analysis, six factors with an eigenvalue greater than one explain 72.2 % of the variance.

Step 5. Interpretation. The oblique rotation solution (Table 3) implies that: the first factor concerns street network design (around 25% variance explained); the second factor relates to urban block diversity (13.1%); the third factor consists of a categorization of nonresidential diversity (9.8%); the fourth factor relates to job accessibility (9.1%) or even job intensity; the fifth factor represents the center accessibility (7.8%); and the last factor can be named as density (6.0%).

Final urban built environment factors
This subchapter develops a FA with land use variables based on data from the Madrid smartphone-based survey where land use variables are calculated with the GPS information of each travel. The results of this case study confirm the 3D’s and accessibility underlying dimensions responsible for land use variables, but it is necessary to make comparisons with prior research in this area. The main comparison findings can be summarized as follows (Table 8).

<table>
<thead>
<tr>
<th>Table 8. Links between previous FA and Madrid case study results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General conclusions of prior research</strong></td>
</tr>
<tr>
<td>Density factor defined at least from population density</td>
</tr>
<tr>
<td>Diversity factor defined at least from land use mix value or entropy index</td>
</tr>
<tr>
<td>Design factor but defined on different ways (street quality factor or average block size, for example)</td>
</tr>
<tr>
<td>Some accessibility factors defined since different points of view (accessibility by car, distance to transit or public transport, distance to CBD, residence/employment in a specialized area, etc.)</td>
</tr>
</tbody>
</table>

Source: de Abreu e Silva et al. (2006); Van Acker and Witlox (2010); de Abreu e Silva and Goulas (2009); Edwing and Cervero (2010); de Abreu e Silva et al. (2012); He and Zhang (2014).

The subchapter reported here is a new step of a the thesis aimed at assessing what and how variables (levels of service, socio-economics, psychological, land use, etc.) influence on travel behavior with introduction of transport policy measures (subchapter 3.1). The contribution of this particular subproject is the definition of land use factors to reach a better understanding of causal relationships between traveler attitudes, socioeconomic characteristics and travel behavior on following studies using different methodologies: structural equations and discrete choice models.
CHAPTER 4: DISCRETE CHOICE MODELS APPROACH
“In general, Discrete Choice Models postulate that the probability of individuals choosing a given option is a function of their socioeconomic characteristics and the relative attractiveness of the option. [...] Discrete Choice Models allow a more flexible representation of the policy variables considered relevant for the study” (Ortúzar and Willumsen, 2011)

Once it has been defined the necessary variables to take into account in a complete study on travel behavior and their relationships, and following Ortúzar and Willumsen (2011) indications, the PhD candidate presents two studies based on DCM to evaluate the transport policy measures:

- **Subchapter 4.1**: a DCM approach with the novelty of two “social influence” variables effect on travel behavior: participation in voluntary activities and receiving help for various tasks.
- **Subchapter 4.2**: a more complete DCM including panel effect (waves 1&2) and built environment variables.

**Subchapter 4.1 – Exploring the role of social capital influence variables on travel behavior**

To include this new step in the thesis, it is necessary to consider some studies that have recognized that users’ social interactions – as well as their perceptions – may influence travel behaviour, especially when changes to the transport system are introduced (Ben-Akiva et al., 2012; Brock and Durlauf, 2003; Carrasco and Miller, 2006). Notwithstanding, key social influence variables are rarely included directly in travel behaviour models possibly because of the difficulty in measuring the degree of integration of people with respect to their spatial proximity and social environment (social capital). However, the social environment may influence travel behaviour and integrating it in transport behaviour models can help to explain why two apparently identical individuals can make completely different decisions when facing an objectively equal situation.

**The relevance of social capital influence**

Social capital variables indicate the social embedding of people in their own contexts (Di Ciommo, 2003; Granovetter, 1985; 2005; Putnam et al., 1994). If the local community is social capital oriented (Bevort, 1994), potential public policies, including transport policies, to be implemented may be slightly different. Recent examples in California (i.e. sustainable community building in Santa Barbara) show how built environment relations (i.e. social capital variables) influence travel behaviour choice. However, social influence needs to fit the needs of people according to their life cycle stage. Hence, there is a scientific interest in including social capital influence variables in a discrete choice model framework (Portney, 2005).

Methodologically, a DCM is commonly used in transport planning applications. However, as discrete choice theory, is fundamentally grounded on individual choices, the treatment of the interdependence of choices among various decision-makers’ and their integration into social networks remains an outstanding challenge (Dugundji, and Walker, 2005). Interestingly, due to the increasing awareness of the influence of social variables in travel behaviour, it now has evidence from a wide range of analyses that individual outcomes are linked to social interaction factors, especially within social groups defined by geographic proximity (Brooks-Gunn, 1993; Glaeser et al., 1992; Bauer and Zimmermann, 1999). In particular, Ostrom (1998) modified the rational utility theory to include social interaction variables, such as reciprocity, trust and reputation, obtaining improved results. Brock and Durlauf (2001) and Skrondal and Rabe-Hesketh
(2003), have already expanded the effort to employ social interaction variables in a structural estimation with binary choice models.

When new infrastructure is introduced, transport planners are especially interested in defining its pro-modal-shift factors. But it is very difficult to know the real causes of a modal shift. Some studies about voluntary travel behaviour change have shown how much travel cost, together with cultural and social values, may influence or even determine traveller’s mode choice (Ampt, 2003). For example, measures for promoting public transport are more effective for new residents and frequent public transport users (Fujii and Taniguchi, 2006). Moreover, attitudes are not necessarily associated with observed behaviour; for example, a person who has a positive attitude to reduce car use cannot change to another mode if s/he lacks information about available alternatives to shift (Eagly and Kulesa, 1997).

In this subchapter it uses data from the first wave of smartphone panel survey of Madrid. Based on current literature investigating the treatment of the interdependence between various decision-makers and its influence on transport modes (Currie and Delbosc, 2010; Dugundji and Gulyas, 2003; Dugundji and Walker, 2006; Deutsch and Goulias, 2013), the questionnaire for the Madrid panel included two specific questions about “social capital influence” for each individual; first, if they received some help for child-care or for housekeeping (helped variable), or if they voluntarily participated in some non-compulsory meetings or activities (voluntary variable). Based on Di Ciommo et al. (2014) it analyses the effect of these “social influences” in explaining the observed shift to metro over and above the classical quantitative attributes typically included in mode choice models (i.e., travel cost, income, trip purpose, gender and car ownership).

To follow the approaches commented above, it incorporates social capital variables directly into a Mixed Logit (ML) model (Train, 2009). As it is explained at subchapter 1.4 (DCM mathematical formulation) the advantage of using a ML is that it allows incorporating random heterogeneity in some key attributes, such as time and cost, and also to consider panel effects.

**Definition of variables and results**

The models of this study consider three transport modes: public transport (PT), walking (W) and Car (which is taken as reference alternative). A set of LOS attributes, socioeconomic variables and other attributes are specified to explain choices. First, two alternative-specific variables: total travel time from origin to destination (Total Travel Time) and personal cost, specified as the ratio between the cost of travel and the available income of the traveller (Personal Travel Cost). The available income is calculated using the average transport consumption expenditure of Spanish households (i.e. 11.6%, see EUROSTAT, 2012) and other social status variables:

\[
\text{Personal Travel Cost}_{ijq} = \frac{\text{travel cost}}{\text{monthly income}} \cdot 0.116 \cdot \frac{\text{household size}}{\text{average monthly trips}}
\]  

Second, three individual-specific variables are included as dummies that did not vary across alternatives: Sex, which takes the value of one for females; car ownership (CO) which takes the value of one for households with one or more cars, and Purpose which takes the value of one if the purpose of the trip is work. Finally, the face-to-face interview questionnaire also included two questions regarding the social capital influences of residents and workers in the catchment area:
- if they participated in non-compulsory meetings or activities (voluntary dummy variable) and
- if they received some help for child-care or housekeeping (helped dummy variable).

Then the model includes the two dummies (Helped and Voluntary) related with the social integration level of the decision-maker. These are two observable characteristics in the sense that they explain a specific situation for each decision-maker. The first is a time consuming variable and the second a time saving variable; thus, both may influence travel behaviour, but probably in a different way.

Table 9 reports the results of two ML models with different specifications, with and without the social interaction effects, in order to test the effect of their inclusion.

### Table 9 Model estimation results (N = 974 observations) (Floridea et al., 2014)

<table>
<thead>
<tr>
<th></th>
<th>ML1</th>
<th>ML2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimates</td>
<td>t-test</td>
</tr>
<tr>
<td>Total Travel Time</td>
<td>-0.036</td>
<td>(-4.9)</td>
</tr>
<tr>
<td>Personal Travel Cost</td>
<td>-0.441</td>
<td>(-6.1)</td>
</tr>
<tr>
<td><strong>Systematic heterogeneity in Travel time</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender: Female</td>
<td>-0.283</td>
<td>(-8.4)</td>
</tr>
<tr>
<td>Car ownership</td>
<td>-0.004</td>
<td>(-7.2)</td>
</tr>
<tr>
<td>Work</td>
<td>-0.003</td>
<td>(-1.9)</td>
</tr>
<tr>
<td>Voluntary</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Helped</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td><strong>Attributes specific for Public Transport</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASC</td>
<td>0.980</td>
<td>(17.0)</td>
</tr>
<tr>
<td>St. Dev. (panel effect)</td>
<td>1.465</td>
<td>(9.1)</td>
</tr>
<tr>
<td>Gender: Female</td>
<td>2.017</td>
<td>(3.3)</td>
</tr>
<tr>
<td>Car ownership</td>
<td>-3.932</td>
<td>(-7.1)</td>
</tr>
<tr>
<td>Work</td>
<td>-0.770</td>
<td>(-1.5)</td>
</tr>
<tr>
<td>Voluntary</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Helped</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td><strong>Attributes specific for Walking</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASC</td>
<td>1.240</td>
<td>(16.9)</td>
</tr>
<tr>
<td>St. Dev. (panel effect)</td>
<td>0.043</td>
<td>(6.1)</td>
</tr>
<tr>
<td>Female</td>
<td>-1.346</td>
<td>(-4.1)</td>
</tr>
<tr>
<td>Car ownership</td>
<td>-5.263</td>
<td>(-5.0)</td>
</tr>
<tr>
<td>Work</td>
<td>-1.438</td>
<td>(-1.9)</td>
</tr>
<tr>
<td>Voluntary</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Helped</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-433.91</td>
<td></td>
</tr>
</tbody>
</table>

As can be seen, all coefficients are significantly different from zero and have the expected sign. A log-likelihood ratio test on ML1 and ML2 confirms that for this particular case study, the latter has higher explanatory power (the models are also clearly superior to their MNL counterparts, not shown in the table). Further, with the inclusion of Voluntary and Helped almost all parameters increase their statistical significance except for Time and Sex, and there seems to be an influence between Time and both social interaction terms. This is expected because a person doing more activities has less time, so the introduction of a voluntary activity should explain part of the story explained by the time variable in ML2. Also, people receiving some help in childcare, should have more free time, and vice versa. On the other
hand, the social attributes also improve the statistical quality of model **ML2**, especially CO and the Work variable related to Walk. In summary, these results confirm that observed behaviour is not only ruled by the typical LOS attributes, but also by a strong social influence component.

The results of model **ML2** imply that the introduction of Voluntary reduces the probability of choosing WALK and PT, with a non-linear influence on the utility function. This social influence variable result could be explained by the effect of activities on mode choice: if somebody participates more in voluntary activities, s/he consumes more time than otherwise and may take into account level of service more (and in peripheral boroughs, such as the case studies examined in this paper, this favours the car mode). In turn, Helped increases the probability of choosing WALK and PT (as when having some help for house and childcare, a person saves some time that can be used for travelling by foot or by public transport).

A breakdown in the analysis of Helped and Voluntary user types (Floridea et al., 2014), confirms the difference between “helped” and “voluntary”. Helped people freely choose to take public transport because they have, on average, more cars and higher income; and voluntary people, on the other hand, have less access to cars and a relatively lower income, and freely decide not to shift to new metro line use. However, the estimation of systematic heterogeneity in travel time for both social variables highlights the same negative sign for both types of users (i.e. they perceive negatively the time restriction). Thus, in this case, the introduction of social capital variables produces counter-intuitive results: car choice is not completely associated with a higher level of income. In fact, Voluntary users carry out more activities and liberate some time by using the car intensively. Their social embedding profile will make them more social active people. However, Helped people who are characterized by a different social embedding, seems more relaxed with respect to time constraints and decide to use their cars less because they have more time since they can use the time of “others”.

**New ways to study social capital influence on travel behavior**

Previous literature has worked under the assumption that the explanatory variables for modal choice (or changes in travel behaviour) are objective (transport cost, travel time, etc.). Recent studies have recognized, however, that users’ social influences – as well as their perceptions, attitudes and awareness – may affect travel behaviour, especially when new infrastructure is introduced (Ben-Akiva et al., 2002; Tudela et al., 2011).

In this study it found that with the inclusion of two such variables (the Voluntary and Helped dummies) the model fit improves significantly and almost all parameters increase their statistical significance (except Time). Apparently, a linear influence exists between travel time and these variables. The main result of the consideration of “helped” and “voluntary” user types is that social variables - as those tested here - could have a positive or negative impact on the use of public transport. It believes that this is a potentially important result for forecasting future public transport demand and to better understand what could be a favourable environment for increasing public transport demand.

In models ML1 and ML2 the Work dummy, related to the PT mode, seems less important for the same reason as above: the direct relationship between different activities (and Voluntary or Helped are two of them). Finally, the Walk mode decreases its probability of choice when the Voluntary variable is introduced, with a non-linear influence on the utility function, but it increases when the Helped variable is introduced. It was expected that both variables could influence travel behaviour in a different way because of their time consuming (Voluntary) or time saving (Helped) effects. Finally, social capital variables can
have an opposite influence on travel behaviour also with respect to an “income variable” and, depending on the social embedness profile, these variables may have a different effect on transport mode use. The smartphone-based panel survey, oriented toward detecting travel mode changes among residents and workers of areas near new metro stations, confirms that around 14.5% of travellers shifted to metro and that the majority of their journeys were characterized as having a work purpose (subchapter 2.2). The consideration of social influence variables improves the explanation of their modal shifts. In particular, the social integration variable Voluntary seems important in determining a specific travel behaviour: 61% of the respondents that shifted to new metro lines are involved in some voluntary activity. Considering the voluntary and helped variables like social interaction indicators, it can say that social influence variables play indeed a role in travel mode choice, at least in the case of Madrid.

**Subchapter 4.2 – Time evolution, social capital and space for exploring travel behavior**

As stated in subchapters above, other studies have also recognized that users’ social interactions—as well as their perceptions—may influence travel behavior, especially when changes to transport systems are introduced (Brock and Durlauf, 2003; Carrasco and Miller, 2006; Ben-Akiva et al., 2012), as well as the land use variables. These studies develop a general framework that extends choice models by including an explicit representation of the environment of the decision maker and some specific elements of his social interaction (i.e. family, friends, and market) on travel choice.

Other studies have shown how built environment relations influence travel behavior, where built environment includes all kind of spatial proximity variables (i.e. design of the site, population density, street pattern, land use mix, distribution of commerce, spatial structure and transportation supply) (Cervero and Kockelman, 1997).

Rich data sources (e.g. multi-day Smartphone travel traces, panel data before-after the implementation of a new transport policy or infrastructure) improve the characterization of travelers and support the forecast of their travel behavior. Actually, a Multi-days Data Survey allows better understanding the dynamics of travel behavior and defining the causality links between variables (i.e. attitudes on travel behavior). There is a growing interest in investigating social capital variables (Galdames et al., 2011; Rieser-Schüssler and Axhausen, 2012; Kamruzzaman et al., 2014).

Nevertheless, key social capital influence variables and design street patterns are rarely included directly in travel behavior models as subchapter above, possibly because of the difficulty in measuring the degree of integration of people with respect to their spatial proximity and the ambiguous effect of social capital attributes.

Following current theoretical approaches, social capital could be a substitute for financial capital, improving the socio-economic status of low-income people, but an additional factor reinforcing the current status of each individual as well, and therefore strengthening the existing inequality (Putnam et al., 1994; Schwanen, 2014). Behind this multifaceted concept, two different social capital variables have been defined: a passive one (receiving help) and an active one (to carry out voluntary activities). Both variables are related with two main aspects of travel choice: time availability and monetary costs. In respect to the time, the “helped” may liberate some additional time for travel, while the second one, “voluntary” activity may create an additional time constraint and generate new social trips (see subchapter 4.1). Regarding
the financial constraints, “helped” represents the additive factor reinforcing the current socio-economic status, while “voluntary” is the investing factor for creating new local social capital.

Recent studies have also shown how all kinds of these spatial proximity variables (i.e. design of the site, population density, street pattern, land use mix, distribution of commerce, spatial structure and transport supply) are part of the built environment relations and influence travel behavior. Since the 1970s the most frequently quoted studies on the impact of the land use patterns on travel behavior have provided significant conclusions. Through a regression analysis, Hurst (1970) demonstrated that higher rates of vehicle trip generation were found among retail and office land uses compared with storage and industrial usages. Newman and Kenworthy (1989) found a significant negative statistical correlation between residential density and transportation-related energy consumption per capita. Urban density is the critical driver of transit ridership (Seskin and Cervero, 1996). The significance of urban density lies in the fact that more people living and/or working in close proximity to transit, the greater the likelihood that the service will be used (Murray et al., 1998). Land-use type and mix also influence transit use, although less than density (Parsons Brinckerhoff, 1996). Land-use mix (diversity) produces a more balanced demand for public transport over time (reducing differences between peak and off-peak periods) and in space (in terms of direction of flow) (Cervero, 2004). Comparative studies of land use and transportation have responded to this problem by promoting neo-traditionalism, with the emphasis on transit-oriented urban design (Loutzenheiser, 1997).

Therefore, at this point of the thesis, it is necessary to develop this background literature including a larger set of variables related to the individual social capital network (i.e. voluntary and helped) (subchapter 4.1); and on the other hand, following Dugundji and Walker (2005) they introduce a large set of land use and street design variables (i.e. centrality variables).

This study incorporates social capital variables, land use and street design variables directly into a Mixed Logit (ML) model. The advantage of using a ML is that it allows incorporating random heterogeneity in some key attributes, such as time and cost, and considering multi-days and panel effects also (Pas, 1998; Cherchi and Cirillo, 2008). Panel data usually offer significant advantages over cross-sectional data as repeated observations from the same individual generally give more precise measurements of changes in individual mobility. This subchapter explores the potential influence on transport mode choice of extrinsic characteristics of individual trip –i.e. time frame using panel data, land use variables using spatial characteristics, social capital variables through “voluntary” and “helped” indicators (subchapter 4.1).

**Formulation of DCM and panel effect**

Using the same modelling framework of subchapter 4.1, the modelling stage started with the estimation of a basic Multinomial Logit (MNL) model, but more general models allow relaxing the three classical restrictions of the MNL model (i.e., absence of random taste variations, unrestricted substitution patterns and correlation of unobserved factors over time, Ortúzar and Willumsem, 2011). In particular, the Mixed Logit (ML) model is completely general in the above senses and simulation of its choice probability is nowadays computationally simple (Train, 2009).

The ML model is, in particular, suitable to account for repeated choices by decision-makers, such as those made by the 250 people who participated in our first wave of panel survey over the course of a week and by 190 respondents of second wave of panel survey. The model specification considered the utility coefficients as varying between individuals but remaining constant through all choice situations for each
person (Train, 2009). To complete the mathematical formulation about DCM presented at subchapter 1.4, the ML probability that accounts for this panel effect takes the following form:

\[ P_{ij} = \int \prod t P_{ijt}(\beta) f(\beta|\theta) d\beta \]  

(22)

where \( P_{ij} \) is the conditional choice probability (it has the same expression as in (2) but it is defined for each repeated choice \( t \)); the parameters \( \beta \) (including the ASC) may vary according to a so called “mixing distribution” \( f(\beta|\theta) \) over the population (\( \theta \) are known as population parameters). Therefore, \( P_{ij} \) is the unconditional probability that individual \( i \) will choose the sequence of alternatives \( j \) along time given a vector of individual-specific coefficients \( \beta \). The probability in (3) is a multi-dimensional integral that can be efficiently estimated through simulated maximum likelihood methods (Ortúzar and Willumsem, 2011).

Different models are estimated for introducing social capital and built environment variables in ML models. Table 10 shows the integration of dummies variables sets (i.e. social capital and built environment variables) and panel data -in a stepped way- ML models estimation.

**Table 10. Estimation of ML models for exploring travel behavior determinants**

<table>
<thead>
<tr>
<th>Estimated Models</th>
<th>#1A: 0 state</th>
<th>#1B: 0 state + Panel Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>#2A: 0 state + Social Effect</td>
<td>#2B: 0 state + Panel Effect + Social Effect</td>
<td></td>
</tr>
<tr>
<td>#3A: 0 state + Land Use Effect</td>
<td>#3B: 0 state + Panel Effect + Land Use Effect</td>
<td></td>
</tr>
<tr>
<td>#4A: 0 state + Social Effect + Land Use Effect</td>
<td>#4B: 0 state + Panel Effect + Social Effect + Land Use Effect</td>
<td></td>
</tr>
</tbody>
</table>

**NOTE:** 0 state (wave 1 with LOS+socioeconomic variables); Panel Effect (waves 1&2); Social Effect (social capital influence variables); and Land Use Effect (environment characteristics).

**Definition of variables and results**

Using the smartphone-based panel survey (two waves), three different transport modes are considered: car, public transport and walking. The ML of this study are distinguished according to combinations of explanatory variables of travel choices: level of service (LOS) attributes (i.e. travel time costs and travel cost), socioeconomic variables, social influence variables and built environment characteristics. All the variables have been used previously:

- **Level of service (LOS) attributes:** Two alternative-specific variables are defined as LOS: total travel time from origin to destination (‘total travel time’) and personal cost, specified as the ratio between the cost of travel and the available income of the traveler (‘personal travel cost’).

- **Social capital variables:** The model includes the known two further dummies related to the level of social integration of the decision-maker/user: ‘voluntary’ (if s/he voluntarily participates in some non-compulsory meeting or activity=1, no=0) and ‘helped’ (if s/he received some help for childcare or housekeeping=1, no=0). The inclusion of “Voluntary” and “Helped” as social capital variables seems to improve the set of current used variables for estimating mode choices by combining both the nature of social network variables in generating social trips with the nature of spatial proximity variables (subchapter 4.1).
• **Built Environment characteristics and street design:** Land use variables have been calculated within the Station Service Area (SSAs) method. Station Service areas were obtained using Geographical Information System (GIS), and are based on distances across the street network. The considered threshold distance was 900 meters, which is the maximum distance that most people (95th percentile) are willing to walk in order to access to the Metro and other Public Transport networks in Madrid (García-Palomares et al., 2013). Obtained SSAs through the GIS have been intersected with various urban variables such as density, diversity and design (Cervero and Kockleman, 1997); and with accessibility dimensions (Geurs and van Wee, 2004). “Population density” and ‘employment density’ have been chosen as density variables for characterizing service area. ‘Population density’ was calculated as inhabitants/hectares and ‘employment density’ as employment/hectares.

Two additional variables of “land use mix” have been used. Firstly, the ratio of employment per inhabitant was computed as ‘job ratio’. Secondly, a more general land use mix variable (i.e. ‘mix’) was measured using the reciprocal of the variation coefficient of the area within SSAs-that includes various land uses- (higher values indicate higher diversity in uses). Both measures are easily computable and interpretable. Fillion (2001) found that mixed-use of suburban centers have been more successful in attaining higher transit use than just residential suburban areas. However, the rest of the different categories of land uses within SSAs have been studied separately to improve the knowledge about the diversity and accessibility dimension: hectares (ha) of commercial and educational (“equipment”), hectares of ‘industry’, hectares of ‘offices’, hectares of ‘infrastructures’ and hectares of ‘parks and recreations’. To measure the “center accessibility” (i.e. accessibility to the center), the distance in kilometers from each SSAs to Center Business District (‘CBD distance’) was also included, as well as ha of brownfield development (‘Brownfield’).

As it is explained at subchapter 3.2, an urban design variable is calculated using the street network layer ‘street density’ within SSAs. Also the “street network design” variables (remoteness centrality, betweenness centrality, straightness centrality and reach centrality) are included. To complete the descriptive analysis of the subchapter 3.2 about these built environment variables, Table 11 shows results about mean, standard deviation, median, skewness and kurtosis for these.
Table 11. Descriptive analysis of Built environment variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Sd</th>
<th>Median</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>'population density'</td>
<td>173.4</td>
<td>97.9</td>
<td>173.2</td>
<td>0.238</td>
<td>-0.770</td>
</tr>
<tr>
<td>'employment density'</td>
<td>79.7</td>
<td>97.0</td>
<td>40.7</td>
<td>1.838</td>
<td>2.798</td>
</tr>
<tr>
<td>'job ratio'</td>
<td>0.45</td>
<td>5.82</td>
<td>0.25</td>
<td>2.11</td>
<td>4.99</td>
</tr>
<tr>
<td>'CBD distance'</td>
<td>9.10</td>
<td>2.01</td>
<td>6.72</td>
<td>2.68</td>
<td>7.94</td>
</tr>
<tr>
<td>'equipment'</td>
<td>630.41</td>
<td>0.17</td>
<td>141.77</td>
<td>3.93</td>
<td>8.34</td>
</tr>
<tr>
<td>'industry'</td>
<td>75.00</td>
<td>0.18</td>
<td>3.10</td>
<td>2.72</td>
<td>7.03</td>
</tr>
<tr>
<td>‘mix’</td>
<td>165.15</td>
<td>137.49</td>
<td>149.21</td>
<td>0.87</td>
<td>-0.31</td>
</tr>
<tr>
<td>‘infrastructures’</td>
<td>763.33</td>
<td>157.40</td>
<td>162.15</td>
<td>3.47</td>
<td>5.17</td>
</tr>
<tr>
<td>‘offices’</td>
<td>40.45</td>
<td>39.85</td>
<td>13.41</td>
<td>2.50</td>
<td>5.89</td>
</tr>
<tr>
<td>‘brownfield’</td>
<td>612.12</td>
<td>171.15</td>
<td>0.00</td>
<td>5.35</td>
<td>8.34</td>
</tr>
<tr>
<td>‘parks and recreations’</td>
<td>526.54</td>
<td>64.44</td>
<td>214.66</td>
<td>2.64</td>
<td>6.61</td>
</tr>
<tr>
<td>‘street density’</td>
<td>162.86</td>
<td>0.50</td>
<td>169.19</td>
<td>-0.46</td>
<td>0.02</td>
</tr>
<tr>
<td>‘remoteness centrality’</td>
<td>0.25</td>
<td>8.89</td>
<td>0.25</td>
<td>0.60</td>
<td>0.19</td>
</tr>
<tr>
<td>‘betweenness centrality’</td>
<td>0.02</td>
<td>2.11</td>
<td>0.01</td>
<td>3.45</td>
<td>1.89</td>
</tr>
<tr>
<td>‘straightness centrality’</td>
<td>0.20</td>
<td>0.20</td>
<td>0.19</td>
<td>1.03</td>
<td>1.34</td>
</tr>
<tr>
<td>‘reach centrality’</td>
<td>0.23</td>
<td>0.03</td>
<td>0.23</td>
<td>0.79</td>
<td>0.67</td>
</tr>
</tbody>
</table>

Table 12 shows the loglikelihood, chi-squared and adjusted rho-squared for the set of estimated ML models. In general, results show that models without panel effect (A) fit worse than models including data of both waves of panel data (B). Moreover, there is an improving statistically significant effect of model fit by adding social capital and built environment variables (A1A, A2A, A3A, and A4A). A first result of comparing various ML models shows that there is not a relevant improving fit between models A4A and A1B.

Table 12. Loglikelihood, Chi-Squared and Adjusted Rho-Squared for Estimated Models

<table>
<thead>
<tr>
<th>Loglikelihood/ Chi-Squared / Adjusted Rho-Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1A: -2284.5 / 3074.4 / 0.44478</td>
</tr>
<tr>
<td>#1B: -2069.4 / 4036.5 / 0.49375</td>
</tr>
<tr>
<td>#2A: -2270.6 / 3102.2 / 0.44971</td>
</tr>
<tr>
<td>#2B: -2044.9 / 4085.5 / 0.49974</td>
</tr>
<tr>
<td>#3A: -2201.5 / 3240.3 / 0.47418</td>
</tr>
<tr>
<td>#3B: -1950.8 / 4273.7 / 0.52276</td>
</tr>
<tr>
<td>#4A: -2180.5 / 3282.4 / 0.48164</td>
</tr>
<tr>
<td>#4B: -1922.1 / 4331.0 / 0.52977</td>
</tr>
</tbody>
</table>

Individuals are considered to have a choice among the three modes of travel. Our study attempted to find out whether better models could be obtained when using social capital and Built Environment variables. Table 13 presents the estimation results of four ML models with different specifications, with and without the social interaction and built environment effects, to test the effects of their inclusion, and with and without panel effect: A1A, A4A, A1B and A4B. For every estimator the respective t-value is indicated in brackets. The ML models are estimated using an own code (available on request from the PhD candidate) written within the R statistics package (Croissant, 2011). The results highlight that the increase of Public transport fare reinforce the significance of income for choosing car instead of Public transport.
Table 13. ML results (N=5562 observations)

<table>
<thead>
<tr>
<th>Parameters</th>
<th>#1A</th>
<th>#4A</th>
<th>#1B</th>
<th>#4B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Travel Time</td>
<td>-0.128 (-18.4)</td>
<td>-0.150 (-15.9)</td>
<td>-0.059 (-18.9)</td>
<td>-0.085 (-20.1)</td>
</tr>
<tr>
<td>Personal Travel Cost</td>
<td>-0.267 (-9.5)</td>
<td>-0.156 (-5.6)</td>
<td>-0.170 (-13.4)</td>
<td>-0.114 (-18.8)</td>
</tr>
<tr>
<td>Systematic heterogeneity in Travel Time</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voluntary</td>
<td>-0.004 (-1.7)</td>
<td>-0.008 (-3.1)</td>
<td>-0.004 (-1.7)</td>
<td>0.011 (3.7)</td>
</tr>
<tr>
<td>Helped</td>
<td>-0.011 (-4.7)</td>
<td>-0.040 (-11.1)</td>
<td>-0.011 (-4.7)</td>
<td>-0.031 (-8.8)</td>
</tr>
<tr>
<td>Attributes specific for Public Transport</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASC</td>
<td>4.827 (8.3)</td>
<td>5.029 (3.7)</td>
<td>4.667 (13.0)</td>
<td>2.889 (3.6)</td>
</tr>
<tr>
<td>St. Dev. (panel effect)</td>
<td>0.084 (15.4)</td>
<td>0.124 (14.3)</td>
<td>0.143 (23.5)</td>
<td>0.170 (21.9)</td>
</tr>
<tr>
<td>Gender: female</td>
<td>-2.010 (-8.7)</td>
<td>-2.004 (-7.0)</td>
<td>-1.715 (-11.5)</td>
<td>-1.119 (-6.4)</td>
</tr>
<tr>
<td>Age</td>
<td>4.825 (4.2)</td>
<td>0.059 (3.9)</td>
<td>0.039 (5.2)</td>
<td>0.070 (6.7)</td>
</tr>
<tr>
<td>Income</td>
<td>-0.0001 (-0.1)</td>
<td>0.0003 (1.9)</td>
<td>-0.0004 (-4.7)</td>
<td>-0.0003 (-2.5)</td>
</tr>
<tr>
<td>Child</td>
<td>-3.223 (-11.3)</td>
<td>-3.744 (-8.9)</td>
<td>-1.412 (-7.7)</td>
<td>-1.774 (-7.2)</td>
</tr>
<tr>
<td>Family</td>
<td>-0.058 (-0.6)</td>
<td>-0.334 (-2.4)</td>
<td>0.139 (2.0)</td>
<td>-0.024 (-0.3)</td>
</tr>
<tr>
<td>Car ownership</td>
<td>-3.499 (-11.9)</td>
<td>-3.414 (-8.3)</td>
<td>-3.693 (-18.0)</td>
<td>-2.748 (-11.6)</td>
</tr>
<tr>
<td>Voluntary</td>
<td>-0.490 (-7.9)</td>
<td>-0.238 (-7.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Helped</td>
<td>2.246 (6.8)</td>
<td>1.010 (4.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population density</td>
<td>0.002 (0.8)</td>
<td>0.007 (4.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employment density</td>
<td>0.013 (3.9)</td>
<td>-0.0001 (-0.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Job ratio</td>
<td>-2.697 (-4.6)</td>
<td>1.254 (5.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mix</td>
<td>-0.002 (-0.4)</td>
<td>0.005 (2.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equipment</td>
<td>-0.0001 (-0.9)</td>
<td>-0.0002 (-3.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Industry</td>
<td>-0.001 (-1.1)</td>
<td>-0.003 (-3.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Offices</td>
<td>0.003 (0.9)</td>
<td>0.005 (3.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parks and recreations</td>
<td>-0.0001 (-1.2)</td>
<td>0.0002 (1.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBD distance</td>
<td>-0.048 (-3.9)</td>
<td>-0.035 (-3.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brownfield</td>
<td>-0.0001 (-1.8)</td>
<td>-0.0008 (-2.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Street density</td>
<td>0.001 (0.2)</td>
<td>-0.012 (-3.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remoteness centrality</td>
<td>-0.048 (-1.1)</td>
<td>-0.093 (-7.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Betweenness centrality</td>
<td>-0.164 (-0.9)</td>
<td>-0.708 (-1.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Straightness centrality</td>
<td>-0.128 (1.7)</td>
<td>0.334 (3.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reach centrality</td>
<td>0.394 (0.4)</td>
<td>0.718 (1.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attributes specific for Walking</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASC</td>
<td>4.736 (7.6)</td>
<td>3.580 (2.5)</td>
<td>4.047 (9.3)</td>
<td>1.336 (2.6)</td>
</tr>
<tr>
<td>St. Dev. (panel effect)</td>
<td>0.825 (18.8)</td>
<td>0.705 (16.9)</td>
<td>0.411 (25.0)</td>
<td>0.569 (30.6)</td>
</tr>
<tr>
<td>Gender: female</td>
<td>-0.700 (-2.8)</td>
<td>-0.656 (-2.1)</td>
<td>-1.146 (-6.3)</td>
<td>-0.556 (-2.7)</td>
</tr>
<tr>
<td>Age</td>
<td>0.063 (2.8)</td>
<td>0.026 (1.5)</td>
<td>0.029 (3.0)</td>
<td>0.038 (3.0)</td>
</tr>
<tr>
<td>Income</td>
<td>0.0003 (3.0)</td>
<td>0.0005 (2.5)</td>
<td>-0.0001 (-0.2)</td>
<td>0.0004 (2.9)</td>
</tr>
<tr>
<td>Child</td>
<td>-2.180 (-6.9)</td>
<td>-2.405 (-5.4)</td>
<td>-1.033 (-4.6)</td>
<td>-1.370 (-4.7)</td>
</tr>
<tr>
<td>Family</td>
<td>-0.231 (-2.1)</td>
<td>-0.434 (-2.9)</td>
<td>-0.016 (-0.2)</td>
<td>-0.117 (-1.2)</td>
</tr>
<tr>
<td>Car ownership</td>
<td>-2.182 (-6.6)</td>
<td>-1.933 (-4.3)</td>
<td>-2.220 (-9.0)</td>
<td>-1.44 (-5.0)</td>
</tr>
<tr>
<td>Voluntary</td>
<td>0.006 (0.2)</td>
<td>0.489 (2.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Helped</td>
<td>2.085 (5.6)</td>
<td>1.430 (5.1)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Assessment Methodology Applied to Demand Measures to Change Urban Mobility Behavior

As can be seen, all coefficients are significantly different from zero and have the expected sign. A log-likelihood ratio test on #1A and #1B, #4A and #4B confirms that for this particular case study, the forth ML models has indeed higher explanatory power. Furthermore, with the inclusion of social capital and built environment variables almost all parameters increase their statistical significance except for Time and Gender. The marginal utility of total travel time seems to be influenced by social interaction terms. This is expected, as a person doing more activities has less time and should reveal a higher value of time. On the other hand, people receiving help (e.g. in childcare) probably needs it and would not have enough time, otherwise, to deal with all their activities (interestingly, although they have relatively higher incomes they tend to choose PT and walk more often, spending more time); this helps explaining why their values of time are even higher than those who do voluntary activities, and therefore, they generally decide to not participate in voluntary activities (Putnam, 2000). This result confirms that observed behavior is not only ruled by the typical LOS attributes, but also by a strong social influence component.

The results of model #4A imply that the introduction of “voluntary” social capital variable reduces the probability of choosing PT, but increase the probability to choose walk mode. Therefore, as shown by Sharmeen and Timmermans (Sharmeen and Timmermans, 2013), these trips are more likely done by private modes (i.e. car, bike and walk) than by public transport. In general, for social trips people choose private modes that present additional degrees of “perceived” freedom with respect to PT (Jakobsson et al., 2000). This result is confirmed and increases its significance in #4B model. Finally, the comparison between ML models that use one wave (i.e. A) or two-waves panel survey (i.e. B) shows that an extensive survey could be sufficient for better estimating travel behavior of users.

### Measures of fit

<table>
<thead>
<tr>
<th>Measure</th>
<th>#1A</th>
<th>#1B</th>
<th>#4A</th>
<th>#4B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-likelihood</td>
<td>-2284.5</td>
<td>-2180.5</td>
<td>-2069.4</td>
<td>-1922.1</td>
</tr>
<tr>
<td>$\chi^2(0)$</td>
<td>0.44</td>
<td>0.48</td>
<td>0.49</td>
<td>0.53</td>
</tr>
</tbody>
</table>

As can be seen, all coefficients are significantly different from zero and have the expected sign. A log-likelihood ratio test on #1A and #1B, #4A and #4B confirms that for this particular case study, the forth ML models has indeed higher explanatory power. Furthermore, with the inclusion of social capital and built environment variables almost all parameters increase their statistical significance except for Time and Gender. The marginal utility of total travel time seems to be influenced by social interaction terms. This is expected, as a person doing more activities has less time and should reveal a higher value of time. On the other hand, people receiving help (e.g. in childcare) probably needs it and would not have enough time, otherwise, to deal with all their activities (interestingly, although they have relatively higher incomes they tend to choose PT and walk more often, spending more time); this helps explaining why their values of time are even higher than those who do voluntary activities, and therefore, they generally decide to not participate in voluntary activities (Putnam, 2000). This result confirms that observed behavior is not only ruled by the typical LOS attributes, but also by a strong social influence component.

The results of model #4A imply that the introduction of “voluntary” social capital variable reduces the probability of choosing PT, but increase the probability to choose walk mode. Therefore, as shown by Sharmeen and Timmermans (Sharmeen and Timmermans, 2013), these trips are more likely done by private modes (i.e. car, bike and walk) than by public transport. In general, for social trips people choose private modes that present additional degrees of “perceived” freedom with respect to PT (Jakobsson et al., 2000). This result is confirmed and increases its significance in #4B model. Finally, the comparison between ML models that use one wave (i.e. A) or two-waves panel survey (i.e. B) shows that an extensive survey could be sufficient for better estimating travel behavior of users.

### Time evolution, social capital and space for exploring travel behavior

The “Voluntary” parameter changes its sign between public transport and walking modal choice becoming positive for walking (see table 13). This result is coherent with those obtained by Sharmeen and Timmermann (2013), who showed the trend of people to choose all kind of private mechanized or not modes, including bike and walking, for social trips. Nevertheless, "helped", the social capital variable that
doesn't generate social trips, keeps the same positive sign between public transport and walking modal choice (see table 14).

The comparison between models without panel effect (A) and models with panel effect (B) shows that both Social capital network variables (i.e. "voluntary" and "helped") improve their explanatory potential between waves #1 and #2. This result highlights the consistency of social capital variables as explanatory variables, at least for the case of Madrid.

Between the first and second panel survey waves, built environment variables (i.e. land use and street design) keep the same explanatory power that is higher for centrality variables for individual modal choice.

To finish this study it is necessary to point out: (1) since high costs in terms of time and money are one of the biggest limitations when building data panels, researchers could evaluate if a new wave of panel data survey is necessary or it will be sufficient to extend the set of questions checking additional explanatory variables. In fact, the integration of social capital variables could improve the model fitness and travel behavior explanation through a one wave smartphone survey. And then (2) the importance of distinguishing social versus spatial influence networks in transportation mode choice behavior can improve the explanation of travel behavior choice. Finally, the social and spatial network embeddedness profile of people may have a different effect on transport mode use. Therefore transport-planners should be aware for future transport policies implementation.
CHAPTER 5:

HYBRID MODEL APPROACH
Coming back to chapter 1, the overall objective of the thesis is to develop a stepped methodology that integrated diverse perspectives to evaluate the willingness to change patterns of urban mobility in Madrid. The diverse perspectives used so far have been based on the following aspects:

- Justifying the treatment of subjectivity in travel behavior (subchapter 2.1).
- Analysis of causal relationships between both objective and subjective personal variables, and travel behavior to capture pro-car and pro-public transport intentions (subchapter 3.1).
- Identifying built environment dimensions on travel behavior (subchapter 3.2).
- Exploring the potential influence of individual trip characteristics and social influence variables on transport mode choice (subchapter 4.1).
- Exploring the potential influence on transport mode choice of extrinsic characteristics of individual trip using panel data, land use variables using spatial characteristics and social influence variables (subchapter 4.2).

Then, the 5th step is how to integrate the results above in a unique way. The methodology proposed by the author is a hybrid DCM model using sequential approach to minor difficulties in its application and interpretation.

This chapter presents the results of a hybrid DCM following Madrid’s database. At the end, findings and main conclusions are discussed (chapter 6.1), practical implications are drawn (subchapter 6.2), and future researches provided (subchapter 6.3).

Introduction

Weis and Axhausen (2009) indicate that the travel behavior observed reflects equilibrium conditions both behaviorally, as well as the underlying network flows and travel times, individuals can adapt their travel behavior after a change in the generalized cost on several levels: (i) the decision to leave home (i.e., to participate in out-of-home activities on a given day); (ii) the adaptation of these activities; (ii) the scheduling of activities; (iv) the choice of locations for carrying out activities (destination choice); and (v) the choice of an origin-destination connection (mode and route choice). Thus, a behavioral DCM model deals the latter two levels, which effectively constitute the second to fourth steps in the classic four step model. The upper levels of the demand generation process (from i to iii above) have been dealt by the literature with an aggregate methodology that handles indirect and multiple relationships between all the variables that influence a travel behavior process (Golob, 2003; Moore et al., 2013): Structural Equation Modeling (SEM). Reviewing many case studies on travel behavior knowledge, it is difficult to have a brainstorm of how to integrate the five levels above to evaluate the effect of transport policy measures to achieve a modal shift. This is the final objective of this thesis, and in fact, of the present study.

Methodologically, in this study factorial and path analyses through a Multiple-Indicator Multiple-Cause (MIMIC) model are used to understand and describe the relationship between the different psychological and environmental constructs with social influence and socioeconomic variables. It is a more general study than the presented in this thesis above. As it was explained at Subchapter 1.4, the MIMIC model generates Latent Variables (LVs) to be incorporated sequentially into Discrete Choice Models (DCM) where the levels of service and cost attributes of travel modes are also included directly to measure the effect of the transport policies that have been introduced in Madrid during the last three years in the context of the economic crisis.
The inclusion of psychological and land-use factors through a Multiple-Indicator Multiple-Cause (MIMIC) following a SEM approach helped to improve the fitness level of DCM and to provide an understanding of the role of Level Of Service (LOS) and cost attributes in the decision making process when new transport policy measures are included (Ben-Akiva et. al, 2002; Raveau et. al, 2010; Yoon and Goulias, 2010). The inclusion of Latent Variables (LVs) in DCM has reemerged as an analysis and discussion topic after losing some of the importance that made it an interesting subject in the 1980s. Furthermore, evidence is mainly based on US data and North-European countries. Therefore, this final study adds some new evidence from a Spanish perspective to the research debate. The data used for this paper are collected from the known two panel smartphone-based survey (n=255 and 190 same respondents, respectively) of Madrid.

Formulation of sequential hybrid discrete choice model
The aim of this final part of the thesis is to study the role of psychological and environmental LVs on the mode choice process in a general way. These factors are measured by mean of psychometric and geographical tools, respectively, which are fit into DCMs through a MIMIC approach. The general formulation of a sequential hybrid DCM and his advantage opposite the simultaneously hybrid form has been justified at the subchapter 1.4.

In sequential estimation the problem is treated in two stages, separating the LV and DCM interactions. First, the MIMIC model is solved to obtain parameter estimators for the equations relating the LVs with the explanatory variables and perception indicators. With these parameters in the structural equation (7), expected values for the LVs of each individual and alternative are obtained. In this way, the LVs can be added to the set of variables of the DCM, and their parameters are estimated together with those of the traditional variables in a second stage.

For the purposes of this study, besides the traditional sequential approach, it considers using a Mixed Logit (ML) to run the DCM, which allows to consider not only the expected value of the LVs, but also its variability (from MIMIC model), which represents a better reproduction of reality. The ML model is in particular suitable to account for repeated choices by individual decision-makers (between week-days or/and different waves). This model specification treats the utility coefficients as varying between individuals but remaining through all choice situations for each person (Train, 2009).

Definition of variables and results
Explanatory variables. Based on the findings of previous studies and the data availability aforementioned, following variables are included as explanatory variables for the MIMIC model: SEX (male = ‘1’; female = ‘0’); AGE; INCOME (monthly household income); CHILD (presence of child = ‘1’, no = ‘0’); FAMILY (number of family members); HELPED (if assisted for child-care or for housekeeping = ‘1’, no = ‘0’); VOLUNTARY (voluntary participation in some non-compulsory meetings or activities = ‘1’, no = ‘0’).

Indicators. For each trip, perceptions towards public transport were measured using 4-point Likert scale to generate a LV for the MIMIC model about the ‘comfort perception’ (subchapter 3.1): ACCESS_TIME, PHYSICAL_ACCESS, TRAVEL_TIME, WAITING_TIME, TRANSFER, SPACE (inside the public transport), SAFETY, and CLEANLINESS. Given that perception towards public transport was asked for each trip, the average perception as arithmetic mean is used assuming that the Likert scale can be treated as a continuous variable (Wang and Law, 2007). The indicators about built environment characteristics within a service area have been calculated within the station service area (SSAs) as the rest of the thesis.
Once the SSAs were defined with a GIS, they have been intersected with various environment indicators that that hypothetically favor transit use: density, diversity and design (Cervero, R. and Kockelman, 1997); and accessibility dimensions (Geurs and van Wee, 2004) for the MIMIC model. At subchapters 3.2 and 4.3 appear a complete description about which land use indicators defined each built environmental LV applying a factor analysis methodology. A list of these indicators is: POP_DEN (population density); EMP_DEN (employment density); JOB_RATIO (ratio of employment per inhabitant); MIX (variation coefficient of the area covered by different land uses within SSAs); EQUIPMENT (hectares (ha) of commercial and educational land use); SINGLE_RES (ha of single-family residential); MULTI_RES (ha of multifamily residential); INDUSTRY (ha of industry); OFFICES (ha of offices); INF (ha of infrastructure that promotes economic activity, such as roads, highways, railroads, airports, electricity, telecommunications, water supply and sanitation); GREEN (ha of parks and recreations); DIST_CBD (distance of each SSAs to Center Business District); BROWNFIELD (ha of land available for building); and four centrality measures: REMOTENESS, BETWEENESS, STRAIGHTNESS and REACH centrality.

**MIMIC model**

In a previous study (subchapter 3.2) some of the MIMIC model measurement relations such as (2) were studied by the PhD candidate using factor analysis to guarantee their correct specification. Using the land use indicators described above, six environmental LVs are justified (‘street network design’, ‘urban block diversity’, ‘nonresidential diversity’, ‘job accessibility’, ‘center accessibility’ and ‘density’); but to capture comfort perceptions of public transport it is necessary to include an additional psychological LV: ‘comfort perception’. Seven explanatory variables are included in the MIMIC model as exogenous variables through structural equations (1): SEX, CHILD, AGE, FAMILY, INCOME, HELPED and VOLUNTARY.

The MIMIC model is fitted using AMOS 20.0 software package (Byrne, 2001). This model fitting is done using a covariance based structural analysis, also referred to as method of moments, consisting in minimizing the difference between actual sample covariances and those implied by the model parameters (Bollen, 1989). In the AMOS software package, computing intercepts for the endogenous variables is only feasible when using the maximum likelihood approach. The model assumes direct causal relationships between certain dependent variables, and thus goes further than merely capturing these relationships via error correlations. For reasons of space, it only presents the details for relationships (measurement and structural equations, Figure 1) of the best model estimated according to the best goodness of fit indices (Schumacker and Lomax, 2004).

Then, based on this MIMIC identification and PhD candidate’s hypotheses from their own previous studies, it can be concluded that this model structure is expected, except for the following two findings: (i) the effect of the ‘street network design’ LV is not significant; (ii) the accepted general recommendation (Anderson and Gerbing, 1988) for running a MIMIC model is to have at least three or four indicators per LV, i.e., ‘center accessibility’ and ‘density’ LVs have been excluded (see subchapter 3.2), so this might imply a flaw in the survey design to fully capture the LVs’ variability. Moreover, to improve the value of standardized regression weights ($\alpha_{lp}$) it is necessary to apply the logarithm for explanatory variables with scale dimension: AGE, FAMILY and INCOME and the inverse of logarithm for the STREET_DEN indicator.
Regarding the structural equations results for all the alternatives, the ‘comfort perception’ towards public transport is measured by using five explanatory variables: SEX ($\alpha_{11}=-0.100$), CHILD ($\alpha_{12}=-0.067$), logINCOME ($\alpha_{15}=0.199$), HELPED ($\alpha_{16}=0.241$) and VOLUNTARY ($\alpha_{17}=-0.260$). This psychological LV is very influenced by the novel social influence variables considered: HELPED in a positive way and VOLUNTARY negatively. The ‘urban diversity’ is measured by SEX ($\alpha_{21}=0.129$) and logAGE ($\alpha_{23}=-0.263$). Higher age is associated with living in an area covered by not many land uses. The other diversity LV that explains the hectares of parks, infrastructures and industries -‘nonresidential diversity’- is measured by three explanatory variables related to family children: CHILD ($\alpha_{32}=0.129$), logFAMILY ($\alpha_{34}=0.082$) and HELPED ($\alpha_{36}=0.114$). Finally, the ‘job accessibility’ is measured negatively by logFAMILY ($\alpha_{44}=0.120$) and by the novel social influence variables: HELPED ($\alpha_{46}=-0.076$) and VOLUNTARY ($\alpha_{47}=-0.072$).

**Hybrid DCM**

Once the expected values of the LVs are calculated for each respondent from the MIMIC model, DCMs are estimated. These models are formulated on the basis of LOS and other socioeconomic variables (subchapter 4.1): ‘travel time’, ‘personal travel cost’, ‘gender=male’, ‘car ownership’, ‘purpose=work’. Moreover, this thesis includes the known (subchapters above) social influence variables (‘voluntary’ and ‘helped’) to evaluate the true importance of this social aspect on a general framework.

Table 15 presents the estimation results. For every estimator the respective t-value follows between parentheses. The models’ log-likelihood at convergence and $r^2$ index (calculated with respect to the constants-only model) are shown as well. In order to evaluate the importance of the LVs inclusion and to facilitate the conclusions, all the DCMs estimated below have a ML model structure with panel correlation accounted for by the ‘travel time’ and ‘personal travel cost’ random parameters.

The first model (ML1) is the simplest specification, as it does not include the effect of the LVs. The LOS parameter signs are correct (consistent with microeconomic theory): lower for car mode. But when PhD candidate have examined the database it is found that many individuals made economically inconsistent
choices, and that this behavior can only be explained through psycho-sociological aspects. For this reason is important to include on the simplest model the social influence variables (‘voluntary’ and ‘helped’): not significant in walking, but the public transport ‘voluntary’ and ‘helped’ constants presents a high significance of its parameters. There is an influence between ‘travel time’ and ‘voluntary’. It is quite trivial, because when I do more activities, I have less time, so the introduction of voluntary activity explains a part of the history explained by time variable. On the other hand, ‘travel time’ and ‘helped’ have also little colinearity because if I have some help or not in child care, I have more or less free time, respectively. At first ‘voluntary’ reduces the probability of choosing public transport, and ‘helped’ increases it. With this result is difficult to know if to be a ‘voluntary’/‘helped’ person implies low/high probability of choosing public transport, respectively; or vice versa. The following model results will solve this hesitation.

Table 15. DCMs results (N=5562 observations)

<table>
<thead>
<tr>
<th>Parameters</th>
<th>ML1 (st.dev)</th>
<th>ML2 (st.dev)</th>
<th>HM1 (st.dev)</th>
<th>HM2 (st.dev)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Travel Time (mean)</td>
<td>-0.074 (27.0)</td>
<td>-0.062 (17.1)</td>
<td>-0.083 (25.2)</td>
<td>-0.067 (16.7)</td>
</tr>
<tr>
<td>Total Travel Time (st.dev)</td>
<td>0.148 (24.9)</td>
<td>0.148 (24.8)</td>
<td>0.115 (23.1)</td>
<td>0.116 (23.0)</td>
</tr>
<tr>
<td>Personal Travel Cost (mean)</td>
<td>-0.142 (12.8)</td>
<td>-0.146 (13.3)</td>
<td>-0.097 (17.1)</td>
<td>-0.047 (4.1)</td>
</tr>
<tr>
<td>Personal Travel Cost (st.dev)</td>
<td>0.385 (25.2)</td>
<td>0.386 (25.0)</td>
<td>0.584 (37.8)</td>
<td>0.472 (24.3)</td>
</tr>
</tbody>
</table>

**Systematic heterogeneity in Travel Time**

| Voluntary | -0.002 (-0.6) | --- | 0.004 (1.5) |
| Helped | -0.015 (-4.5) | --- | -0.043 (-10.9) |

**Attributes specific for Public Transport**

| ASC | 5.516 (22.5) | 5.363 (21.2) | 4.806 (12.9) | 6.200 (15.6) |
| Gender: male | 1.673 (11.7) | 1.641 (11.5) | 1.389 (9.2) | 1.098 (7.2) |
| Car ownership | -4.424 (-22.8) | -4.409 (-22.7) | -3.487 (-18.6) | -3.347 (-17.6) |
| Purpose: Work | 0.262 (2.0) | 0.284 (2.1) | 0.313 (2.3) | 0.157 (1.2) |
| Voluntary | -0.943 (-7.0) | -0.976 (-6.1) | -1.255 (-8.8) | -0.894 (-5.3) |
| Helped | 0.420 (3.0) | 0.652 (3.8) | 0.787 (5.2) | 0.800 (4.4) |
| Comfort perception | --- | --- | -0.408 (-5.4) | -0.746 (-10.0) |
| Urban diversity | --- | --- | -6.653 (-2.1) | -7.815 (-2.5) |
| Nonresidential diversity | --- | --- | -0.005 (-5.6) | -0.006 (-5.7) |
| Job accessibility | --- | --- | 0.030 (6.9) | 0.019 (4.4) |

**Attributes specific for Walking**

| ASC | 5.351 (19.4) | 5.179 (17.9) | 3.370 (7.1) | 4.586 (9.6) |
| Gender: male | 1.105 (6.1) | 1.078 (5.9) | 1.058 (5.6) | 0.690 (3.6) |
| Car ownership | -2.699 (-11.8) | -2.687 (-11.7) | -1.699 (-7.3) | -1.795 (-7.7) |
| Purpose: Work | -0.692 (-4.3) | -0.882 (-5.2) | -0.827 (-4.6) | -0.846 (-4.9) |
| Voluntary | -0.263 (-2.6) | -0.319 (-1.6) | -0.434 (-2.5) | 0.033 (0.2) |
| Helped | -0.045 (-0.3) | 0.248 (1.2) | 0.432 (2.4) | 0.495 (2.9) |
| Comfort perception | --- | --- | 0.181 (1.7) | -0.034 (-0.3) |
| Urban diversity | --- | --- | -0.113 (-2.9) | -0.116 (-3.0) |
| Nonresidential diversity | --- | --- | -0.004 (-3.3) | -0.004 (-3.7) |
| Job accessibility | --- | --- | 0.027 (5.0) | 0.007 (1.4) |

**Measures of fit**

| Log-likelihood | -2068.1 | -2067.0 | -1971.4 | -1959.4 |
| L2(0) | 0.495 | 0.496 | 0.519 | 0.522 |

*Significant at 95% level of confidence

Other study with the same database (subchapter 4.1) justifies several interactions among some respondents’ attributes and decision makers’ characteristics (e.g. travel time), that allow accounting for systematic heterogeneity in the individual preference. The second model (ML2) includes this effect, but only on social influence variables (‘voluntary’ and ‘helped’) to facilitate the comparisons. Due to the colinearity between social influence variables and ‘total travel time’, the constant associate with travel time reduces its statistical significance. The systematic heterogeneity (i.e. taste) in travel time for ‘helped’ is negative. By that is meant that if I have some help for house and child care, I have some time that I can
use for going to work by public transport; not for utilitarian reasons but I have no choice. The rest of the parameters do not change much their specifications.

The third model (HM1) is based on ML1, but is a hybrid model including the four LVs resulting from the MIMIC model above. When psychological and land use factors are added to the discrete choice model framework, models improve their fitness and statistical significance (Yoon and Goulias, 2010; Moore et al., 2013). With the inclusion of these LVs, it can be seen that mode specific constants (ASC) reduce their relative importance and statistical significance, and that LVs are statistically significant (except ‘comfort perception’ towards public transport for walking users). Apparently, this is showing that the observed behavior is not only ruled by the maximum utility criterion, but also by a strong psychological-environment concept, that develops without the mediation of cognitive processes during choice. It is probable that this fact is hidden behind the modal constants in the ML1 and ML2 models. Car users have more positive ‘comfort perception’ towards public transport to the users of bus, metro, etc. This statement corroborates the previous conclusion about having people who go to work by public transport not for utilitarian reasons, but sometimes they have no choice. Regarding environmental LVs, the more diversity place of residence is, the less probability to use of public transport or walking. On the other hand, public transport and walking choices increase with the ‘job accessibility’. The four LVs are measured by, among others, ‘gender’ variable, decreasing its significance in this hybrid model. And again, the social influence variables are significant (even walking mode). LR-test comparing ML1 with HM1 improves from \(-2068.1\) to \(-1971.4\).

Model performance increases when LVs are properly included. Indeed, according to the LR-test last model (HM2), which allows for systematic taste variations in travel time for the social influence variables, is easily the best fit. It must be noted that, except travel time parameter, variables with the highest t-values are precisely those associated with a psycho-sociological construct: systematic heterogeneity in travel time for ‘helped’ respondents and ‘comfort perception’ for public transport users. As it occurred at ML2, the taste variation for ‘voluntary’ variable is not significant; and the environmental importance during mode choice is not affected.

Testing the effect of psychological and built environmental latent variables

Using as a starting point the principle that the choice of a transport mode is the key role played by public transport planners in policy making (Ortúzar and Willumsen, 2011), this study develops a sequential hybrid DCM methodology to analyze the effect of the transport policies introduced in Madrid during the last four years, in the context of the world economic crisis. This final study of this thesis does not only to provide a framework on how to integrate some research topics already developed in previous studies and emerging methods on travel behavior, when assessing the introduction of any transport policy, but also to add new evidences from a Spanish perspective to the research debate on how to change patterns of urban mobility. The data used comes from the two known panel smartphone-based survey of Madrid.

On the basis of this analysis of an urban modal choice case, it is possible to affirm that hybrid DCMs are clearly superior in fit to an equivalent DCM without incorporating LVs. The LVs approach used in the hybrid DCMs is an advance over the typical dummy variables approach because it does not depend on the subjective definition of boundaries for generation of the values of 0 to 1 required in this last case. The consideration of LVs in a DCM allows capturing the real importance that the LOS and cost variables have on the individual decision-making process.
The signs of LOS attributes are consistent with microeconomic theory: time and cost parameters represent a disutility to individuals. Respect of environmental LVs included, higher diversity places of residence (higher mixture of land uses and street density) are associated with more car use. On the other hand, public transport and walking choices increase with places close to job accessibility. Regarding the psychological perception towards public transport, car users have more positive perception to the users of public transport. Finally, with the inclusion for systematic taste variations in travel time for two novel social influence variables (voluntary and helped), models improved their fitness and statistical significance. It is recommended that future research uses a survey design to fully capture all the LVs’ variability on travel behavior; not only other environmental (density and design) and psychological factors (e.g. expectancy and norms), but also other social influence aspects (e.g. number of friends at Facebook, Twitter and LinkedIn) very correlated with the free time to travel.

The results confirm previous findings of this thesis (subchapters 3.1, 3.2, 4.1 and 4.2) – in a more general framework, certainly- on the theory that transport policy actions are more likely to be effective when car use has been first disrupted, because: (1) in areas with higher job accessibility, public transport is more likely to be used since the use of car is more restricted; (2) many people using public transport on their way to work do not do it for utilitarian reasons, but because no other choice is available; and (3) the more things to do for a whole day (household responsibilities, voluntary activities, etcetera), the more use of car.
CHAPTER 6:
CONCLUSIONS AND RECOMMENDATIONS
Subchapter 6.1 – Discussion of results and contributions

Automobiles are popular mainly because of their flexibility: they allow us to decide when and where we travel and to control our micro-environments while we do so. However, car-based travels produce numerous negative externalities. To reduce the use of car, most transport policies focus on involve creating new public transport infrastructures or congestion tolls and parking fees without a strong analysis about the population’s attitudes and the resulting intentions towards transport choice. Because of the complexities associated with attempting to change patterns of urban mobility, different aspects of that willingness were studied in this thesis, i.e.: subjective variables (subchapter 2.1); the mediation treatment of car ownership variable to examine causal relationships in travel behavior (subchapter 3.1); factors which represent the built environmental dimension (subchapter 3.2); social influence variables that improve the explanation of the modal shift (subchapter 4.1); panel effect to evaluate accurately the change in the transportation system (subchapter 4.2); psychological and built environment latent variables in a discrete choice model to capture the real importance that the level-of-service and cost variables have on the individual decision-making process (chapter 5).

Theoretical contributions

As stated in the introduction, recent studies have also recognized that users’ social interactions –as well as their perceptions– may influence travel behavior, especially when changes to transport systems are introduced (Ben-Akiva et al., 2012). Past transport research has led to an increased understanding of behavioral processes; but, key social influence variables are rarely included directly in travel behavior models because of the difficulty in measuring the degree of integration of people with respect to their spatial proximity and social environment (social capital influence). On the other hand, the influence of land use patterns on travel behavior has been subject of many previous studies (Meurs and Haaijer, 2001; Kockelman, 1997), but links between international and Spanish perspectives are rarely dealt with. This thesis has solved many of those questions.

The theoretical contributions of this thesis are focused on:

- Data gathered from a novelty Revealed Preferences survey of the New Station Users (RP-NSU) allows the explanation of the modal shift by considering inertia effect (introduced by, e.g., Cherchi and Manca, 2011) as one variable explaining the individual’s utility function regarding the modal choice. Hence by simplifying the general theory of the inertia effect, a positive relationship is found between the pre-opening disutility alternative (positive inertia) and the likelihood of changing to a new and unknown transport mode.

- The causal effects of socioeconomic and comfort-perception variables on pro-public transport intention are much less than the causal effects of pro-car intention on pro-public transport intention. Moreover, pro-car intention is mainly influenced by car ownership.

- The centrality measures to spatial systems: remoteness, betweenness, straightness and reach (developed by Ravulaparthy and Goulias, 2014) represent the “street network design” factor (introduced by, e.g., Magnanti and Wong, 1984; Marshall and Garrick, 2011) under a novel perspective with satisfactory results on travel behavior studies.

- The consideration of social influence variables improves the explanation of the modal shift. In particular, the social integration variable voluntary seems important in determining a specific travel behavior: 61% of the respondents that shift to new metro lines are involved in some
voluntary activity. Considering the voluntary and helped variables like social interaction indicators, it can say that social influence variables indeed play a role in travel mode choice, at least in the case of Madrid.

- The importance of distinguishing social versus spatial influence networks in transportation mode choice behavior improves the travel behavior explanation.
- The observed behavior is not only ruled by the maximum utility criterion, but also by a strong psychological-environment concept, that develops without the mediation of cognitive processes during choice.
- Regarding environmental latent variables, the more diversity the place of residence, the more difficult the use of public transport or walking. On the other hand, public transport and walking choices increase with the ‘job accessibility’ (or job intensity).
- Many people using public transport on their way to work do not do it for utilitarian reasons, but because no other choice is available.

Methodological contributions

Starting from the principle that the choice of a transport mode is the key role played by public transport planners in policy making (Ortúzar and Willumsen, 2011), traditionally travel behavior has been studied from the utilitarian point of view. These studies focused on instrumental motivations ignored those psychological aspects that may play a relevant role. These studies were used to analyze transport demand, based on origin/destination surveys together with Discrete Choice Model (DCM) built on the Random Utility Maximization theory (RUM), which offered an experimental approach to the problem of modeling in a discrete choice context (McFadden, 1974; McFadden, 1981; Train, 2009; Ortúzar and Willumsem, 2011). Despite the growing interest in activity-based modeling within transportation research, there was a need for more knowledge about the causal relationship in order to determine how the urban layout influences travel behavior.

Then, the methodological contributions of this thesis are focused in:

- The incorporation of perception questions in the novelty RP-NSU for application in a DCM –even with a simplified survey like the one used at subchapter 2.1- improves the fitness and statistical quality of a DCM that directly measures modal shift through the utility of the pre-opening alternatives.
- This thesis proposes the use of technology such as GPS (already often integrated in smartphones) to track individual movements. This allows measuring with high precision important characteristics of the trip such as origin/destination and travel times. Moreover, information is available in real time, all the movements are recorded avoiding the typical errors of omitting short trips or overestimating travel times. The use of this technology alleviates also the task of the respondents, as they do not have to remember or take notes of all the trips made.
- To illustrate the use of SEM in travel behavior research to model the influence of independent variables on dependent variables, the author converts the travel behavior variable into two new variables that take account of modal choice (pro-public transport and pro-car intentions) and behavior with satisfactory results.
- The bootstrapping method highlights the key importance of the mediation treatment of “car ownership” variable in SEM to examine causal relationships in travel behavior, as the relationship between “household responsibilities” and pro-car intention has a total effect –including the “car ownership” variable– of 0.388, but an indirect effect of 0.172 (subchapter 3.1).
Assessment Methodology Applied to Demand Measures to Change Urban Mobility Behavior

- FA is the most common data reduction technique. Factors derived through such data reduction techniques are also referred to as latent variables. Although, as the name latent implies, these variables are not observable, certain effects on measurable (manifest) variables can be observed (Srinivasan, 2001). Thus, this thesis justifies that FA methods can assess and explain the structure in a set of correlated and observed built environmental variables in terms of a small number of latent variables or factors (subchapter 3.2).
- When examining the database it was found that many individuals made “economically inconsistent” choices. This behavior can only be explained through psycho-sociological aspects and is hidden behind the modal constants ASC in DCM without social influence effects. With the inclusion of the voluntary and helped dummies almost all parameters increased their statistical significance except time (subchapter 4.1).
- The integration of social capital influence variables could improve the model fitness and travel behavior explanation at a survey lower cost (subchapter 4.2).
- Hybrid DCMs are clearly superior in fit to an equivalent DCM without incorporating latent variables, i.e., when psychological and land use factors are added to the DCM framework, models improve their fitness and statistical significance (chapter 5).

**Subchapter 6.2 – Policy implications**

Since travel behavior is performed in a physical and social context, changes in urban mobility may be difficult even with a strong motivation to reduce car use. For example, the possibilities for using public transport may be poor and expectancies from family and friends may encourage the use of private car. As suggested by several researchers (e.g. Diekmann and Preisendörfer, 2003), the context may establish boundaries when attitudinal factors influence travel behavior. However, different transport policy measures may be used to remove various barriers hindering pro-environmental travel behavior (Eriksson et al., 2008). For example, public transport may be improved to increase the possibilities for traveling more sustainably, and the cost for using the car may be increased in order to interrupt habitual car use (Fujii and Gärling, 2007) or to enhance the motivation to reduce car use (Jakobsson et al., 2002).

The results of this thesis demonstrate the importance of car users’ evaluation of policy measures in order to understand the reactions toward policy. These results indicate that reasons for why policy measures are ineffective and/or unacceptable may be found by examining car user’s policy specific beliefs. In consequence, the following policy recommendations can be drawn:

- One of the main reasons why modal shift from car is so difficult is the perceived travel time. An urban trip by car is always perceived as taking less time than on public transport. Therefore, by way of example, in the case of new metro infrastructures the stress should be put not only on the speed of the trains.
- Regarding to the RP-NSU, the pre-opening car users’ perceived cost savings in a short-medium trip by metro, and show less resistance to the modal shift from private to public transport. Nevertheless, since car users only perceive fuel costs—not captured by the RP-NSU—, a policy question worth to be addressed should be to make them aware of the real nature of the costs of car use (depreciation, insurance, taxes, maintenance and inspections).
- Commute trips (work and studies) are more likely to produce a modal shift than other trip purposes. This paves the way for green transport plans (packages of practical measures to reduce car use for journeys to and from work and for business travel), as an effective and
sustainable response to increase public travel demand on routes with limited or non-existent public transport services (López-Lambas and Aparicio, 2004).

- Transport policy actions are more likely to be effective when pro-car intention has first been disrupted because: (1) cars provide places where people can feel a positive “comfort” in contrast to public transport; and (2) the more “household responsibilities”, the greater the pro-car intention.
- Information campaigns and new or improved public transport infrastructures are unlikely to be sufficient to encourage people to consider an alternative mode of transport that increases new pro-public transport intention, as “household responsibilities” directly impact pro-car intention and indirectly impact pro-public transport intention.
- The main result of the consideration of “helped” and “voluntary” types of user in a model to study the willingness to change patterns on travel behavior, is that social variables as those tested here could have a positive or negative impact on the use of public transport. This is potentially a key result for forecasting future public transport demand and to better understand what could be a favorable environment for increasing public transport demand.
- Regarding the psychological perception towards public transport, car users have a more positive perception than the users of public transport, at least in the case of Madrid.

Summarizing, transport policy actions are more likely to be effective when car use has been first disrupted, because: (1) in areas with higher job intensity, public transport is more likely to be used since the use of car is more restricted; (2) many people using public transport on their way to work do not do it for utilitarian reasons, but because no other choice is available; and (3) the more things to do for a whole day (household responsibilities, voluntary activities, etcetera), the more use of car.

Based on the results obtained, at least three strategies to increase the modal shift in favor of more environmental friendly transport systems and to reduce inertia effects in changing mode choice, can be suggested, e.g.: (1) increase the areas where car use is restricted; (2) raise congestion charges; and (3) promote hybrid car use for the pro-car intention users.

**Subchapter 6.3 – Future researches**

The many difficulties associated with changing travel behavior call for continued research within this area. Despite the results of this thesis are in line with previous studies and tend to corroborate - from a broader perspective- and extend the results found in studies carried out in Spain and in other European countries, the author proposes some new researches:

- The influence of habit (leading to inertia) in the choice process has been largely discussed in the literature (Gärling and Axhausen, 2003), but with different interpretations (see review on this issue by Bamberg, and Schmidt, 2003). The data used in this thesis comes from the HABIT project (Habit and Inertia in mode choice behavior: a data panel for Madrid) with the methodological idea of incorporating inertia in DCMs. For economic and practical implications, it was only possible to collect mobility data from two waves; but since each wave provides data for five consecutive days, after each wave it would be also possible to test innovative models to account for variability among days of the week as well as several types of structural dependence, both true and spurious. Innovative models –with mobility data from four or five waves- would consist in the formulation and estimation of dynamic DCMs. This would allow studying the complex problem of inertia or habit, which cannot be approached with the traditional models.
• Regarding GPS data, although the technology used was very good, there are still some problems such as capturing correctly the signal, especially in urban areas, which requires studying an accurate methodology: (i) to improve the measurement precision and the strength of the signal, to avoid errors in the tracks; (ii) to increase the number of coordinates (seconds between signals) reducing the charge consumption; (iii) to use appropriated algorithms to align typical zigzag erroneous measures; (iv) to use map-matching algorithms to match GPS signals with the network implemented in GIS and finally (v) to automate the passage between the information collected into a the dataset ready to be used for the mathematical analyses.

• It is recommended that future research uses a survey design to fully capture all the LVs’ variability on travel behavior; not only environmental (density and design) and psychological factors (e.g. expectancy and norms), but also other social influence aspects (e.g. number of friends at Facebook, Twitter and LinkedIn) very correlated with the free time to travel.

• Activity-based travel modeling has begun to make significant progress towards a more behavioral framework for simulating household travel behavior and understanding. These new generations of activity-based travel models are based on new neighborhood scale transportation planning policies; requiring more detailed information on household demographics and employment characteristics and the location of activities. Therefore, if it is necessary to approach the willingness to change patterns of urban mobility with the introduction of transport policy measures in vicinity, the use of DCM approach is not enough.

• One important field of research should be to address problems of evaluation of policy impacts and issues of equity. Predominantly economic evaluation techniques need to be complemented by multicriteria methods capable of measuring non-monetary aspects of mobility and neighborhood, environmental quality and their distribution across privileged, and disadvantaged socioeconomic and spatial groups.

• The stepped methodology described in this thesis is too aggregate in substance, space and time to match the sophistication of contemporary travel demand models and to respond to the requirements in spatial resolution of neighborhood-scale spatial policies to promote public transport, cycling and walking as well as the named activity-based travel modeling. These deficiencies suggest the agenda for modeling research. Future land use and transportation models will need to be more disaggregate, more integrated and more responsive to environmental issues. In fact, this thesis has not found significant relationships between travel behavior variables and the day-to-day variability (defined by Pas, 1983) due to a limited treatment of this variable, because of the insufficient nonworking trips registered.

• This thesis proposes an overall methodology to evaluate the introduction of some transport policy measures together: five new metro stations, a 25% average public transport fare rise, one hour extension of the on-street parking frame time, new fleets and stations, real-time information at most metro/train stops, some bus stops and on-board most metro/train and buses, traffic calming (30 km/h or less) at most residential streets, etc. Since these policy measures (known as Travel Demand Measures (TDM) should be effective in reducing travel demand, detailed knowledge on the behavioral adaptations (e.g., reduced car use, increased use of alternative travel modes) made in response to different measures separately is needed for a successful implementation of a single transport policy measure. A more detailed specification of how TDM measures influence travel choices separately was proposed, i.e., by Eriksson (2008).


international colloquium on the behavioural foundations of integrated land use and transportation models, Toronto.


Schwanen T. et al. (2014). Rethinking the links between social and transport disadvantage through the lens of social capital. Transportation Research Part A: Policy and Practice, in press.


Assessment Methodology Applied to Demand Measures to Change Urban Mobility Behavior


Zumkeller, D. (2009). The Dynamics of Change – latest results from the German Mobility Panel. 12th International Conference on Travel Behaviour Research, Jaipur, India.
ANNEX
Annex – Mobility survey (smartphone-based panel survey)

This annex includes the questionnaire of the smartphone-based panel survey that was used in this thesis. Phd candidate and his supervisors are very grateful to Floridea Di Ciommo, Juan de Dios Ortúzar and Elisabetta Cherchi for their support in preparing the questionnaire. The following table summarizes the basic data of the two waves that are based on the questionnaire.

<table>
<thead>
<tr>
<th></th>
<th>Wave 1</th>
<th>Wave 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survey start date</td>
<td>September 2011</td>
<td>September 2012</td>
</tr>
<tr>
<td>Survey close date</td>
<td>June 2012</td>
<td>May 2013</td>
</tr>
<tr>
<td>Survey pattern</td>
<td>5 week days</td>
<td>5 week days</td>
</tr>
<tr>
<td>Nº of the respondents</td>
<td>255</td>
<td>190</td>
</tr>
</tbody>
</table>

Then, it presents the cited questionnaire –in Spanish- which was programmed through an Android® application by G@TV-UPM. The smartphone-based survey (Figure 2a) for the first wave (September 2011) had two main phases. The first consisted of a face-to-face interview to gather personal data on the respondent (PARTE A). In the second phase, respondents were given the smartphone and asked to register the daily trips they made during the five workdays (PARTE B) and other confidential questions (i.e., income) to answer at any moment along the week (PARTE C).

PARTE A: DURANTE LA ENTREGA DE LOS SMARTPHONES

1. Sexo
   1.a. Hombre
   1.b. Mujer

2. Situación en el hogar
   2.a. Padre
   2.b. Madre
   2.c. Hijo/a
   2.d. Cónyuge o pareja sin hijos
   2.e. Otras situaciones

3. ¿Dispone de permiso de conducir válido? Desde cuándo?
   3.a. Sí
   3.b. No

4. ¿Dispone de permiso de moto válido? Desde cuándo?
   4.a. Sí
   4.b. No

5. ¿Dispone de coche?
   5.a. Sí
   5.a.1. Tipo de coche
   5.a.1.a. <2 litros
   5.a.1.b. =2 litros
   5.a.1.c. >2 litros
   5.a.2. Año de matrícula
   5.a.3. Tipo de combustible
   5.a.3.a. Gasolina
   5.a.3.b. Diesel
   5.a.3.c. Otros

   5. b. No

6. ¿Dispone de moto?
   6.a. Sí
6.a.1. Tipo de moto
   6.a.1.a. <250 cm³
   6.a.1.b. =250 cm³
   6.a.1.c. >250 cm³

6.a.2. Año de matrícula

6.a.3. Tipo de combustible (Se repiten las respuestas del punto 5.a.3.)

6.b. No

7. ¿Dispone de bicicleta?
   7.a. Sí
   7.b. No

8. ¿Dispone de abono de transporte?
   8.a. Sí
   8.a.1. Tipo de abono
      8.a.1.a. A
      8.a.1.b. B1
      8.a.1.c. B2
      8.a.1.d. B3
      8.a.1.e. C1
      8.a.1.f. C2
      8.a.1.g. E1
      8.a.1.h. E2
   8.a.2. ¿Cuántos viajes hace con el abono durante los días laborables?
   8.a.3. ¿Quién paga el abono?
      8.a.3.a. El encuestado
      8.a.3.b. El empleador
      8.a.3.c. Otros

8.b. No

9. Por su trabajo: ¿se ve obligado a utilizar el coche?
   9.a. Sí
   9.b. No

10. ¿Participa usted en alguna reunión de barrio, de vecinos o de trabajo, aunque no sea obligatoria su asistencia?
   10.a. Sí
   10.b. No

11. Si necesita ayuda para diversas tareas (cuidado de niños, del hogar, etc.), tiene personas que puedan echarle una mano?
   11.a. Sí
   11.b. No

**PARTE B: AL FINAL DE CADA VIAJE (CUANDO DAN AL STOP)**

10.a.1. Motivo del Viaje
   10.a.1.a. Trabajo
   10.a.1.b. Estudios
   10.a.1.c. Médico
   10.a.1.d. Ocio
   10.a.1.e. Compras
   10.a.1.f. Otros

10.a.2. Lugar de salida
   10.a.2.a. Casa
   10.a.2.b. Trabajo
   10.a.2.c. Otro (dirección) ___________

10.a.3. Lugar de llegada
   10.a.3.a. Casa
10.a.3.b. Trabajo
10.a.3.c. Otro (dirección) __________

10.a.4. Modo de transporte
10.a.4.a. A pie
   10.a.4.a.1. ¿Cuánto tiempo ha tardado en realizar el viaje?
   10.a.4.a.2. ¿Tendría la posibilidad de realizar este viaje en coche?
   10.a.4.a.2.a. Sí
      10.a.4.a.2.a.1. ¿Tendría la posibilidad de aparcar?
         10.a.4.a.2.a.1.a. Sí
         10.a.4.a.2.a.1.b. No
   10.a.4.a.2.b. No
   10.a.4.a.3. ¿Tendría la posibilidad de realizar este viaje en moto?
      (Si ha respondido 6.b. no se hace esta pregunta)
   10.a.4.a.3.a. Sí
   10.a.4.a.3.b. No

10.a.4.a.4. Con anterioridad a la subida de la tarifa en el Transporte Público, ¿realizaba este viaje?
   10.a.4.a.4.a. Sí
      10.a.4.a.4.a.1. La subida de la tarifa en el Transporte Público, ¿cambió el modo de transporte del viaje actual?
         10.a.4.a.2.a.1.a. Sí
         10.a.4.a.2.a.1.b. No
   10.a.4.a.4.b. No

10.a.4.b. Bicicleta
   10.a.4.b.1. ¿Tendría la posibilidad de realizar este viaje en coche?
      (Se repiten las preguntas del punto 10.a.4.a.3.
   10.a.4.b.2. ¿Tendría la posibilidad de realizar este viaje en moto?
      (Si ha respondido 4.b. no se hace esta pregunta)
   10.a.4.b.4.a. Sí
   10.a.4.b.4.b. No
   10.a.4.b.3. Con anterioridad a la subida de la tarifa en el Transporte Público, ¿realizaba este viaje? (Se repiten las respuestas del punto 10.a.4.a.4.)

10.a.4.c. Taxi
   10.a.4.c.1. Al inicio del viaje: ¿cuántos metros aproximados caminó para encontrar el taxi?
      10.a.4.c.1.a. <100m
      10.a.4.d.1.b. 100-200m
      10.a.4.d.1.c. 200-500m
      10.a.4.d.1.d. >500m
10.a.4.c.2. Al inicio del viaje: ¿cuánto tiempo tardó en caminar para llegar al taxi?
10.a.4.d.2.a. 1-2min
10.a.4.d.2.b. 2-5min
10.a.4.d.2.c. 5-10min
10.a.4.d.1.d. >10min

10.a.4.c.3. ¿Tendría la posibilidad de realizar este viaje en coche?
(Se repiten las preguntas del punto 10.a.4.a.2.)
10.a.4.c.4. ¿Tendría la posibilidad de realizar este viaje en moto?
(Si ha respondido 4.b. no se hace esta pregunta)
10.a.4.c.4.a. Sí
10.a.4.c.4.b. No

10.a.4.c.5. Con anterioridad a la subida de la tarifa en el Transporte Público, ¿realizaba este viaje? (Se repiten las respuestas del punto 10.a.4.a.4.)
10.a.4.c.6. Coste aproximado del viaje (euros)________
10.a.4.c.7. ¿Quién paga el viaje?
10.a.4.c.5.a. El encuestado
10.a.4.c.5.b. El empleador
10.a.4.c.5.c. Otro

10.a.4.d. Coche (conductor)
10.a.4.d.1. Al inicio del viaje: ¿cuántos metros aproximados caminó para llegar al vehículo?
10.a.4.d.1.a. <100m
10.a.4.d.1.b. 100-200m
10.a.4.d.1.c. 200-500m
10.a.4.d.1.d. >500m

10.a.4.d.2. Al inicio del viaje: ¿cuánto tiempo tardó en caminar para llegar al vehículo?
10.a.4.d.2.a. 1-2min
10.a.4.d.2.b. 2-5min
10.a.4.d.2.c. 5-10min
10.a.4.d.1.d. >10min

10.a.4.d.3. ¿Dónde aparca?
10.a.4.d.3.a. Parking reservado
10.a.4.d.3.b. Parking público
10.a.4.d.3.c. Calzada-Zona SER
10.a.4.d.3.d. Calzada-sin Zona SER
10.a.4.d.3.e. Otros

10.a.4.d.4. ¿Tendría la posibilidad de realizar este viaje en moto?
(Si ha respondido 4.b. no se hace esta pregunta)
10.a.4.d.4.a. Sí
10.a.4.d.4.b. No

10.a.4.d.5. Con anterioridad a la subida de la tarifa en el Transporte Público, ¿realizaba este viaje? (Se repiten las respuestas del punto 10.a.4.a.4.)
10.a.4.d.6. Coste aproximado del viaje (euros)_______
10.a.4.d.7. ¿Quién paga el viaje? (Se repiten las respuestas del punto 10.a.4.c.7.)

10.a.4.e. Coche (acompañante)
10.a.4.e.1. Al inicio del viaje: ¿cuántos metros aproximados caminó para llegar al vehículo? (Se repiten las respuestas del punto 10.a.4.d.1.)
10.a.4.e.2. Al inicio del viaje: ¿cuánto tiempo tardó en caminar para llegar al vehículo? (Se repiten las respuestas del punto 10.a.4.d.2.)
10.a.4.e.3. ¿Tendría la posibilidad de realizar este viaje en coche propio? (Se repiten las preguntas del punto 10.a.4.a.2.
10.a.4.e.4. ¿Tendría la posibilidad de realizar este viaje en moto? (Sí ha respondido 4.b. no se hace esta pregunta)
   10.a.4.e.4.a. Sí
   10.a.4.e.4.b. No
10.a.4.e.5. Con anterioridad a la subida de la tarifa en el Transporte Público, ¿realizaba este viaje? (Se repiten las respuestas del punto 10.a.4.a.4.)
   10.a.4.e.6. Coste aproximado del viaje (euros)________
   10.a.4.e.7. ¿Quién paga el viaje? (Se repiten las respuestas del punto 10.a.4.c.7.)

10.a.4.f. Moto
10.a.4.f.1. Al inicio del viaje: ¿cuántos metros aproximados caminó para llegar al vehículo? (Se repiten las respuestas del punto 10.a.4.d.1.)
10.a.4.f.2. Al inicio del viaje: ¿cuánto tiempo tardó en caminar para llegar al vehículo? (Se repiten las respuestas del punto 10.a.4.d.2.)
10.a.4.f.3. ¿Tendría la posibilidad de realizar este viaje en coche?
(Se repiten las preguntas del punto 10.a.4.a.2.)

10.a.4.f.4. Con anterioridad a la subida de la tarifa en el Transporte Público, ¿realizaba este viaje? (Se repiten las respuestas del punto 10.a.4.a.4.)
   10.a.4.f.5. Coste aproximado del viaje (euros)________
   10.a.4.f.6. ¿Quién paga el viaje? (Se repiten las respuestas del punto 10.a.4.c.7.)

10.a.4.g. Metro
10.a.4.g.1. Número de transbordos
   10.a.4.g.1.a. 0
   10.a.4.g.1.a. 1
   10.a.4.g.1.b. 2
   10.a.4.g.1.c. 3
10.a.4.g.2. Enumere las estaciones donde usted ha realizado el transbordo________ (no hacer esta pregunta si ha contestado 10.a.4.g.1.a.)
10.a.4.g.3. Estación de entrada________
10.a.4.g.4. Estación de salida________
10.a.4.g.5. Al inicio del viaje: ¿cuántos metros aproximados caminó para llegar a la estación del metro? (Se repiten las respuestas del punto 10.a.4.d.1.)
10.a.4.g.6. Al inicio del viaje: ¿cuánto tiempo tardó en caminar para llegar a la estación del metro? (Se repiten las respuestas del punto 10.a.4.d.2.)
10.a.4.g.7. ¿Tendría la posibilidad de realizar este viaje en coche?
(Se repiten las preguntas del punto 10.a.4.a.2.)
10.a.4.g.8. ¿Tendría la posibilidad de realizar este viaje en moto?  
(Si ha respondido 4.b. no se hace esta pregunta)  
10.a.4.g.8.a. Sí  
10.a.4.g.8.b. No

10.a.4.g.9. Con anterioridad a la subida de la tarifa en el Transporte Público, ¿realizaba este viaje? (Se repiten las respuestas del punto 10.a.4.a.4.)  
10.a.4.g.10. Coste aproximado del viaje (euros) (Si ha respondido 6.a. no se hace esta pregunta) ________  
10.a.4.g.11. ¿Quién paga el viaje? (Se repiten las respuestas del punto 10.a.4.c.7.)  
10.a.4.g.12. Evalúe la calidad del Transporte Público escogido con respecto a antes de la abertura de las nuevas estaciones de metro para los siguientes aspectos:   
(1=Peor, 2=Mejorable, 3=Bueno, 4=Muy bueno)  
10.a.4.g.12.1. Tiempo de acceso al Transporte Público  
10.a.4.g.12.1.a 1 (Peor)  
10.a.4.g.12.1.b 2 (Mejorable)  
10.a.4.g.12.1.c 3 (Bueno)  
10.a.4.g.12.1.d 4 (Muy bueno)

10.a.4.g.12.2. Accesibilidad física (Se repiten las respuestas del punto 10.a.4.g.12.1.)  
10.a.4.g.12.2. Accesibilidad monetaria (Se repiten las respuestas del punto 10.a.4.g.12.1.)

10.a.4.g.12.3. Tiempo de viaje  
10.a.4.g.12.4. Tiempo de espera  
10.a.4.g.12.5. Facilidad en el intercambio  
10.a.4.g.12.6. Espacio disponible  
10.a.4.g.12.7. Seguridad  
10.a.4.g.12.8. Condiciones (limpieza, mantenimiento)  

10.a.4.h. Bus  
10.a.4.h.1. Número de transbordos  
10.a.4.h.1.a. 0  
10.a.4.h.1.a. 1  
10.a.4.h.1.b. 2  
10.a.4.h.1.c. 3  
10.a.4.g.2 Enumere las paradas donde usted ha realizado el transbordo. (no hacer esta pregunta si ha contestado 10.a.4.h.1.a.)

10.a.4.h.3. Parada de entrada  
10.a.4.h.4. Parada de salida  
10.a.4.h.5. Al inicio del viaje: ¿cuántos metros aproximados caminó para llegar a la parada de bus? (Se repiten las respuestas del punto 10.a.4.d.1.)
10.a.4.h.6. Al inicio del viaje: ¿cuánto tiempo tardó en caminar para llegar a la parada de bus? (Se repiten las respuestas del punto 10.a.4.d.2.)
10.a.4.h.7. ¿Tendría la posibilidad de realizar este viaje en coche? (Se repiten las preguntas del punto 10.a.4.a.2.)
10.a.4.h.8. ¿Tendría la posibilidad de realizar este viaje en moto? (Si ha respondido 4.b. no se hace esta pregunta)
10.a.4.h.8.a. Sí
10.a.4.h.8.b. No
10.a.4.h.9. Con anterioridad a la subida de la tarifa en el Transporte Público, ¿realizaba este viaje? (Se repiten las respuestas del punto 10.a.4.a.4.)
10.a.4.h.10. Coste aproximado del viaje (euros) (Si ha respondido 6.a. no se hace esta pregunta) _______
10.a.4.h.11. ¿Quién paga el viaje? (Si ha respondido 8.a. no se hace esta pregunta) (Se repiten las respuestas del punto 10.a.4.c.7.)
10.a.4.h.12. Evalúe la calidad del Transporte Público escogido con respecto a antes de la abertura de las nuevas estaciones de metro para los siguientes aspectos (1=Peor, 2=Mejorable, 3=Bueno, 4=Muy bueno) (Se repiten las preguntas del punto 10.a.4.g.12.)

10.a.4.i. Cercanías
10.a.4.i.1.a. 0
10.a.4.i.1.a. 1
10.a.4.i.1.b. 2
10.a.4.i.1.c. 3
10.a.4.i.2. Enumere las estaciones donde usted ha realizado el transbordo________ (no hacer esta pregunta si ha contestado 10.a.4.i.1.a.)
10.a.4.i.3. Estación de entrada________
10.a.4.i.4. Estación de salida________
10.a.4.i.5. Al inicio del viaje: ¿cuántos metros aproximados caminó para llegar a la estación de cercanías? (Se repiten las respuestas del punto 10.a.4.d.1.)
10.a.4.i.6. Al inicio del viaje: ¿cuánto tiempo tardó en caminar para llegar a la estación de cercanías? (Se repiten las respuestas del punto 10.a.4.d.2.)
10.a.4.i.7. ¿Tendría la posibilidad de realizar este viaje en coche? (Se repiten las preguntas del punto 10.a.4.a.2.)
10.a.4.i.8. ¿Tendría la posibilidad de realizar este viaje en moto? (Si ha respondido 4.b. no se hace esta pregunta)
10.a.4.i.8.a. Sí
10.a.4.i.8.b. No
10.a.4.i.9. Con anterioridad a la subida de la tarifa en el Transporte Público, ¿realizaba este viaje?? (Se repiten las respuestas del punto 10.a.4.a.4.)
10.a.4.i.10. Coste aproximado del viaje (euros) (Si ha respondido 6.a. no se hace esta pregunta) _______
10.a.4.i.11. ¿Quién paga el viaje? (Si ha respondido 8.a. no se hace esta pregunta) (Se repiten las respuestas del punto 10.a.4.c.7.)
10.a.4.h.12. Evalúe la calidad del Transporte Público escogido con respecto a antes de la abertura de las nuevas estaciones de metro
para los siguientes aspectos (1=Peor, 2=Mejorable, 3=Bueno, 4=Muy bueno) (Se repiten las preguntas del punto 10.a.4.g.12.)

10.a.4.j. Transporte combinado
10.a.4.j.1. ¿Qué modo de transporte utilizó primero?
  10.a.4.j.1.a Bicicleta (Se repiten las preguntas del punto 10.a.4.b.)
  10.a.4.j.1.b Taxi (Se repiten las preguntas del punto 10.a.4.c.)
  10.a.4.j.1.c Coche (conductor) (Se repiten las preguntas del punto 10.a.4.d.)
  10.a.4.j.1.d Coche (acompañante) (Se repiten las preguntas del punto 10.a.4.e.)
  10.a.4.j.1.e Moto (Se repiten las preguntas del punto 10.a.4.f.)
  10.a.4.j.1.f Metro (Se repiten las preguntas del punto 10.a.4.g.)
  10.a.4.j.1.g Bus (Se repiten las preguntas del punto 10.a.4.h.)
  10.a.4.j.1.h Cercanías (Se repiten las preguntas del punto 10.a.4.i.)

10.a.4.j.2. ¿Qué modo de transporte utilizó después? (Se repiten las respuestas del punto 10.a.4.j.1.)

10.a.4.j.3. ¿Utilizó algún otro modo de transporte?
  10.a.4.j.3.a Sí
    10.a.4.j.3.a.1. ¿Cuál? (Se repiten las respuestas del punto 10.a.4.j.1.)
  10.a.4.j.3.b No

PARTE C: PESTAÑA A PARTE DE LA APLICACIÓN (a rellenar en cualquier momento por el propio encuestado o, en el peor de los casos, a la entrega del aparato)

11. Edad
   11.a. <20
   11.b. 20-24
   11.c. 25-34
   11.d. 35-44
   11.e. 45-59
   11.f. 60-65
   11.g. >65

12. N° de personas que viven en el hogar
   12.a. 1
   12.b. 2
   12.c. 3
   12.d. 4
   12.e. 5
   12.f. 6
   12.g. >6

13. Situación profesional
   13.a. Trabajador fijo
   13.b. Trabajador temporal
   13.c. Trabajador eventual
   13.d. Otras
14. Rama de actividad
   14.a. Educación
   14.b. Sanidad
   14.c. Administraciones públicas, defensa, servicios sociales
   14.d. Otros servicios

15. Ingresos mensuales medios en el hogar
   15.a. <1000€
   15.b. 1000-2000€
   15.c. 2000-3000€
   15.d. 3000-4000€
   15.e. >4000€

16. Aportación personal (en porcentaje) al ingreso del hogar
GLOSSARY
This glossary is a list of specialised terms used in the thesis with which a reader are not expected to be familiar, each with its definition as understood in the text:

- **ADF**: Asymptotically Distribution Free
- **BNL**: Binomial Logit
- **CFA**: Confirmatory Factor Analysis
- **CNL**: Cross-Nested Logit
- **CRTM**: Consorcio Regional de Transportes de Madrid
- **DCM**: Discrete Choice Model
- **EDM**: Encuesta Domiciliaria de Madrid
- **EFA**: Exploratory Factor Analysis
- **EMTA**: European Metropolitan Transport Authorities
- **EMT**: Empresa Municipal de Transportes (Madrid)
- **FA**: Factor Analysis
- **GDP**: Gross Domestic Product
- **GEV**: Generalized Extreme Value
- **GIS**: Geographical Information System
- **GLS**: Generalized Least Squares
- **HABIT**: Habit and inertia in mode choice behavior: a data panel for Madrid (project)
- **HM**: Hybrid Model
- **IIA**: Independence of Irrelevant Alternatives
- **IID**: Identically Distributed
- **INE**: Instituto Nacional de Estadística
- **LOS**: Level Of Service
- **LV**: Latent Variable
- **MCA**: Multiple Centrality Assessment
- **MIMIC**: Multiple Indicator Multiple-Cause
- **MMA**: Madrid Metropolitan Area
- **ML**: Mixed Logit
- **MNL**: Multinomial Logit
- **NL**: Nested Logit
- **OGEV**: Ordered Generalized Extreme Value
- **PCA**: Principal Component Analysis
- **PCL**: Paired Combinational Logit
- **PROCAR**: Pro Car intention
- **PROPT**: Pro Public Transport intention
- **PSTP**: Pudget Sound Transportation Panel
- **PT**: Public Transport
- **RP**: Revealed Preference
- **RP-NSU**: Revealed Preferences survey of New Station Users
- **RUM**: Random Utility Maximization
- **SEM**: Structural Equation Modeling
- **SSA**: Station Service Area
- **TDM**: Travel Demand Measure
- **ULS**: Unweighted Least Squares
- **WLS**: Weighted Least Squares