



# SELECTED PROCEEDINGS

## RESIDENTIAL SELF SELECTION AND GEOGRAPHICAL SCALES IN TRIP FREQUENCY

LISSY LA PAIX, UNIVERSITY OF TWENTER, L.C.LAPAIXPUELLO@UTWENTE.NL  
ANDRÉS MONZÓN, UNIVERSIDAD POLITÉCNICA DE MADRID, ANDRES.MONZON@UPM.ES  
ELISABETTA CHERCHI, TECHNICAL UNIVERSITY OF DENMARK, ELICH@TRANSPORT.DTU.DK  
MANUEL LORO, UNIVERSIDAD POLITÉCNICA DE MADRID, MANUEL.LORO@UPM.ES

This is an abridged version of the paper presented at the conference. The full version is being submitted elsewhere.  
Details on the full paper can be obtained from the author.

ISBN: 978-85-285-0232-9

13th World Conference  
on Transport Research

[www.wctr2013rio.com](http://www.wctr2013rio.com)

15-18  
JULY  
2013  
Rio de Janeiro, Brazil

unicast

# RESIDENTIAL SELF-SELECTION AND GEOGRAPHICAL SCALES IN TRIP FREQUENCY

*Lissy La Paix, Post-doctoral Researcher at Centre for Transport Studies, University of Twente. P.O217-7500 AE Enschede, The Netherlands, [l.c.lapaixpuello@utwente.nl](mailto:l.c.lapaixpuello@utwente.nl)*

*Andrés Monzón, Professor of Transport. Civil Eng., Universidad Politécnica de Madrid, Av. Profesor Aranguren, 28040 Madrid, [andres.monzon@upm.es](mailto:andres.monzon@upm.es)*

*Elisabetta Cherchi, Associate Professor, Department of Transport, Technical University of Denmark, Bygningstorvet 116 Vest, 2800 Kgs. Lyngby, [elich@transport.dtu.dk](mailto:elich@transport.dtu.dk)*

*Manuel Loro, Researcher at Transport Research Centre TRANSyT, Universidad Politécnica de Madrid, Av. Profesor Aranguren, 28040 Madrid, [manuel.loro@upm.es](mailto:manuel.loro@upm.es)*

## ABSTRACT

Neighbourhood representation and scale used to measure the built environment have been treated in many ways. However, it is anything but clear what representation of neighbourhood is the most feasible in the existing literature. This paper presents an exhaustive analysis of built environment attributes through three spatial scales. For this purpose multiple data sources are integrated, and a set of 943 observations is analysed. This paper simultaneously analyses the influence of two methodological issues in the study of the relationship between built environment and travel behaviour: (1) detailed representation of neighbourhood by testing different spatial scales; (2) the influence of unobserved individual sensitivity to built environment attributes.

The results show that different spatial scales of built environment attributes produce different results. Hence, it is important to produce local and regional transport measures, according to geographical scale. Additionally, the results show significant sensitivity to built environment attributes depending on place of residence. This effect, called residential sorting, acquire different magnitudes depending on the geographical scale used to measure the built environment attributes. Spatial scales risk to the stability of model results. Hence, transportation modellers and planners must take into account both effects of self-selection and spatial scales.

## INTRODUCTION

The relationship between built environment (BE) and travel behaviour (TB) has been approached from several points of view along the time. Substantial improvements on the representation of BE have taken place in the past decades. However, there are, at least, five critical points in the existing background about the relationship:

1. Relative importance between BE attributes and socioeconomic (SE) characteristics: This is one of the major problems discussed in the past literature to explain trip frequency. Early studies found that the total number of trips is largely determined by demographic and socio-economic factors, but it is not strongly associated with land use characteristics (Hanson 1982; Kitamura, Mokhtarian et al. 1997). By contrast, other studies found that land use patterns affect trip frequency and layout of urban areas can assist in reducing travel (Ecotec 1993; Agyemang-Duah, Anderson et al. 1995). There are some reasons for these contradictory results: the effect depends on the type of journey studied (work, shopping, all and so on) and on the type of (BE) variables included in an analysis.
2. Systematic comparison: there is no systematic comparison between different dimensions of BE and TB. Trip frequency has been the major focus also in the past decade, but authors mainly refer only to specific purposes or to specific category of people.
3. Neighborhood representation: Frequently, neighborhood type is represented by dummy variables. However, the major problem with neighborhood representation is *cumulative effect* of the BE attributes, which occurs when only one variable is used to represent neighborhood characteristics. Hence the effect of the different factors cannot be disentangled.
4. Geographical scale: this means the spatial levels used to measure BE. The majority of the studies use predefined spatial units based on census tracts, zip codes, or Traffic Area Zones (TAZ) as operational replacement for neighborhoods because urban form data is more available and easily matched to travel data at these scales. Unfortunately the effect of the geographical scale has not received sufficient attentions in the literature and there is no evidence of which geographical scale is the most feasible.
5. Self-selection: A crucial issue when models are used is that they assume that relation has a unique defined direction. As discussed in Brownstone (Brownstone 2008) individuals and/or families choose where to live and work based, among other things, on their preferences for different types and durations of travel. This tendency is called *self-selection*, the existence of which must be accounted for in the study of the relationship BE-TB, if we want to be able to produce valid estimates of the impact of land use policies on behaviour. Cao et al. (Cao, Mokhtarian et al. 2009) provide a review of many empirical studies where self-selection is accounted. In the empirical context, many authors have verified the effect of self-selection regarding to different aspects : Handy *et al.* (Handy, Cao et al. 2005) used a quasi-longitudinal design to investigate the relationship between neighbourhood characteristics and TB while taking into account the role of travel preferences and neighbourhood preferences in explaining this relationship. Bhat and Guo (Bhat and Guo 2007) analyzed car ownership and residential location. Similarly, Pinjari et al. (Pinjari, Pendyala et al. 2007) studied commute mode choice and residential location. More recently, Pinjari et al (Pinjari, Bhat et al. 2009) focused on residential sorting and activity time-use choices considering a comprehensive set of activity-travel environment variables. In the same year, Salon et al (2009) analyzed walk trips, car-ownership and residential location.

However, there is still a gap in the study of self-selection at different scales of BE measures.

Among these five, a discussion about the critical points first and second critical points can be found in La Paix *et al.* (*La Paix, Monzón et al. ; La Paix 2012*). Third, fourth and fifth critical points are analyzed in the present paper: neighborhood type representation, geographical scale and self-selection. For this purpose, trip frequency is estimated through sociodemographic and BE attributes. In this line, this paper includes an analysis of the effect of unobserved factors related to individual sensitivity when faced with BE components

Therefore, analysing the research gap in this field, the goal of the present paper is twofold:

- To test the differences in transport demand models at different geographical scale of neighbourhood
- Verify the effect of preferences in residential location choice and trip frequency, via self-selection test.

This paper is structured as follows: firstly, a literature review of BE representations , geographical scales and self-selection in the context of residential location is included with the aim to emphasise the most relevant methodological contributions of this paper. Following this, description of the case study, model specification and results are included. Finally, the paper concludes with the summary of main findings and future research.

## **CHALLENGES AND LIMITATIONS OF DIFFERENT BUILT ENVIRONMENT REPRESENTATIONS**

It is difficult to define and measure the attributes that characterize BE, because often the several dimensions are generally correlated among them. In general, two approaches are available: aggregated and disaggregated representation of BE. In this section, we firstly describe aggregated representation, called generic representation of neighbourhood type. And secondly, disaggregated representation is approached, which is called “detailed” representation. This is divided in three tenets: density, shape and accessibility. Finally, the most important elements of geographical scale to measure BE are described.

### **Generic Representation**

In the past literature, authors basically described BE as classification of areas, in terms of several characteristics, such as: date of built, i.e. traditional, suburban or early modern (Friedman, Gordon *et al.* 1994); size of the city (Dieleman, Dijst *et al.* 2002); distance from CBD (Bhat and Srinivasan 2005); mobility patterns, i.e. transit or automobile oriented (Brownstone 2008) and income level (Paez, Scott *et al.* 2007; Farber and Paez 2009) . However, the problems with dummy variables are: firstly, effect of neighbourhood characteristics cannot be disentangled; secondly, considering only a single BE factor may exaggerate its effect on TB (Stead and Marshall 2001). Lastly but not less important, BE can have a different impact depending on whether BE measures are tested individually or in combination (Ewing, DeAnna *et al.* 1996).

### **Detailed Representation**

For simplicity, we can explain the “detailed” representation in three tenets of attributes related to three widely studied concepts:

*Density: Several density measures can be mentioned to demonstrate the capability of these kind of measures to represent the influence of neighbourhood type variations on travel patterns, such as: Population density (Frank and Pivo 1994; Cervero and Gorham 1995; Boarnet and Crane 2001; Ewing, Pendall et al. 2003); Employment density (Bhat 1999; Ewing, Pendall et al. 2003); and percentage of workers, either in general or by age cohort (García-Palomares 2009). However, measures related to economic development are not very included in the representation of neighborhood type. The density variables included in this paper represent land-use diversity (commercial, residential and industrial). We include indicators related to economic development by municipality (percentage of workers and gross domestic product). Measures area normalized by squared kilometers, which allows to represent relative position of each municipality depending on the area. Some authors have used instrumental variables (IV) (Boarnet and Sarmiento 1998; Boarnet and Crane 2001) and two stage regression model to estimate population density (Khattak and Rodriguez 2005). The problem with IV arises at the estimation process when one works with instrumental variables and discrete choice models. There are control functions and related approaches today to deal with the case of endogenous “explanatory” variables in the context of discrete choice and other non-linear models (Berry 1995; Louviere 2005). However, these methods need rather intensive computations to identify the error component in the predicted value of the endogenous BE attributes to obtain the correct standard errors in the main equation of interest.*

- Shape: it can be thought as geographical form of specific characteristic. From the concept of Shape can be applied from smaller to larger units of urban form: street, neighborhood or city. Similar to the present paper, other studies measured the street geometry within a quarter-mile of each person’s residence (Frank and Pivo 1994; Cervero and Gorham 1995; Boarnet and Crane 2001; Ewing, Pendall et al. 2003)
- Accessibility: This concept is used in number of scientific fields such as transport planning, geography, urban planning, and transport policy. Thus, there are several methods and indicators to represent accessibility. Different components of accessibility can be identified from a number of definitions and practical measures: land-use components, transportations, temporal and individual (Geurs and Van Wee 2004). Both land-use and transportation components are strongly associated with the representation of BE. Hence, land-use diversity and transit service are carefully designed through Geographical Information Systems (GIS). Transit ridership in Madrid was recently forecasted (Gutiérrez et al. 2011) based on a combination of GIS, multiple regression models and distance-decay functions. They predicted the number of passengers boarding at each station in Madrid Metro network as function of, among other things, land-use mix, employment characteristics and street density. Similarly, García-Palomares (2010) focused on urban sprawl and mobility to work in Madrid. Both studies come up with important correlations between accessibility measures, spatial patterns and trip frequency. Additionally, the paper opens a gap in:

13<sup>th</sup> WCTR, July 15-18, 2010 – Río, Brazil

- 1) Integration of more comprehensive accessibility measures based in GIS;
- 2) Development of more explicit mathematical representation of total trip frequency in Madrid, such as discrete choice models.

## Geographical scale of neighbourhood

Significant progress has been made in the operationalization of neighbourhood concept. Administrative units are used in the majority of studies. However, perception of neighbourhood involves psychological and intangible factors not interpretable in the “predefined” administrative divisions. Spatial units such as census block group or tract data can be too large to capture the variations in urban forms that occur at a much smaller scale. It is recently argued, not only in transportation context, that representations of spatial concepts in statistical models should be based upon the individuals, the place and the problem under study.

Coulton *et al.* (2001) examined the resident’s perception and mental maps, and found discrepancies between resident-defined neighbourhood and census geography. Among this, it is relatively unclear how individuals perceive the “neighbourhood” space and scale, and how they filter spatial information when making spatial choice decisions (Golledge and Gärling 2003; Krizek 2003; Guo and Bhat 2007). Frank and Pivo (1994) a district level data; Boarnet and Sarmiento (1998) and Boarnet and Crane (2001) used three levels of spatial detail, even for not all variables implied in the study: quarter-mile circle, blocks and zip code. However, census and administrative units do not always represent the zonal characteristics at origin and destination points. Additionally, variables measured from different geographical scales were introduced in the same model. This work proved that modelling approach and geographical scale both matter. They found that division by postal code showed the relevance of smaller scales.

Guo and Bhat (2007) illustrated the concept of neighbourhood by examining the effect three operational units of neighbourhood representations (0.4 km, 1.6 km, and 3.2 km as the radii and band size) in household location choice. They found that measuring a set of variables in different spatial units could end in ambiguous results. Similarly, Frank *et al.* (2008) calculated land-use variables for a one kilometre area buffer, based on the distance that can be covered in about a ten-minute walk. The study, among other findings, demonstrates potential benefits of other spatial units out of administrative zones. In the same way, Krizek (2003) states the necessity of understanding different styles of neighbourhood design. The author highlights the importance to capture urban form at a scale sensitive to walking behaviour (e.g. one quarter mile).

As can be concluded, several and different factors influence the relationship between TB and BE. These factors depend, of course, on which dimension (or characteristic) of the TB is considered and on how neighbourhood and geographical scale is defined. The issue of geographical scale has received less attention in the past. Hence, transport researchers and planners should measure neighbourhood characteristics what matters to people over the area that really matters to people (Guo and Bhat 2007). Thus, in order to contribute to the operationalization of neighbourhood representation, this paper explores three geographical scales: buffer, district and municipality, with the aim to verify its effect of different in the study demand models. Additionally, the present paper verifies the influence of those scales in the

analysis of individual's sensitivity to neighbourhood attributes, called residential self-selection. According to this, the next section describes the most relevant findings and needs on this field.

## Residential Self Selection

The relationship between BE and TB could be considered as causal, instead of associative. Self-selection does not mean nonexistent causal relationship. In practice, both causal and associative relationship must be present. "Self-selection" in this context can be described as the association between residential location choice and travel patterns.

The importance of analyzing residential self-selection is because it may confound the association between BE and TB and, as a consequence, it could produce biased results. As stated by Næss (2009), if households were able to self-select this would not mean the BE did not influence TB. On the contrary, the BE enables households to self-select.

Most studies have employed multivariate analysis and accounted for the sorting effect of socio-economic characteristics (Abreu e Silva, Golob et al. 1977; Kitamura, Mokhtarian et al. 1997; Van Acker 2007)); while others focus on the issue of attitude induced self-selection, where they provide a review of many empirical studies where self-selection is accounted (Mokhtarian and Cao 2008),.

The presence of self-selection is suitably accounted in longitudinal approaches than cross-sectional approaches (Krzizek 2003). Handy *et. al* (2005) used a quasi-longitudinal design to investigate the relationship between neighbourhood characteristics and TB while taking into account the role of travel preferences and neighbourhood preferences in explaining this relationship. However, longitudinal studies are not often available, due to expensive costs of data collection and respondents recruitment.

Researchers, have been mainly focus on commute mode choices, related to residential location (Pinjari, Pendyala et al. 2007). More specifically, studies about car-use and its relationship with self-selection are very frequent. Salon (2009) analysed walk trips, car-ownership and residential location. Bhat and Guo (2007) analysed car ownership and residential location. In more general dimensions, Naess (2009) explained total travel on weekdays, in relation to attitudes towards car-use and environment. The study highlighted the role of residential self-selection could have been made more precise if specific attitudes (cycling and walking) had been explicitly included in the models that explain the proportion travelled by bicycle and foot.

However there is still a gap in the research about geographical scales used to capture the effect of self-selection. Hence, the main concern in self-selection analysis is the lack of variation in collected data.

As stated in results of Bohte (2010) the best results are obtained if the action, target, context and time are measured with the same degree of specificity or generality. Hence, to describe BE at the scale that really matter to individuals, play a key role.

Another methodological issue is the method used to test self-selection. There are several methods. Direct questioning, statistical control, instrumental variables, sample selection, propensity score, joint discrete choice models, structural equation models, mutually-dependent discrete choice models, and longitudinal designs. Cao et al. (2009) provide a

review of many empirical studies where self-selection is accounted for. A summary of these studies is reported below with the aim of defending the method used in the context of this paper.

This paper examines the effect of residential self-selection in trip-frequency. According to this, two methods can be considered for accounting self-selection in this context: Joint mixed ordered model and joint discrete choice model. A joint discrete choice model applied in Salon (2009) consisted of estimating a three-tiered nested logit model of residential choice, auto ownership, and walking level. She estimated both simple and full joint models and computed the elasticity regarding population density. In this case, self-selection is the difference between the elasticity of walking level and those conditional on residential location. However, the correlation captured in nested logit models does not mean causality. By contrast, the model formulation in Bhat and Guo (2007) takes the form of a joint mixed ordered response structure that controls for the self-selection of individuals into neighbourhoods based on car ownership preferences holding for both demographic characteristics and unobserved household factors. Similarly, Pinjari et al (2009) studied residential self-selection in activity time-use behaviour via a joint model system. This model accommodates for differential sensitivity to the activity travel environment attributes (ATE) and additionally, the system controls for the self-selection of individuals into neighbourhood-related attributes. The study involves a set of comprehensive variables. Moreover, the joint model demonstrates strong capability to capture both unobserved and observed factors. , additionally, as stated by the authors, this research can be extended to more disaggregate spatial scales.

Hence, the present paper develops a joint mixed ordered because joint models simultaneously accommodate the differential sensitivity to the explanatory attributes and choices. Additionally, the model explicitly represents the temporal constraints of these two decision processes faced by each individual.

According to the findings in the literature review, the objective of this paper is to deal simultaneously with two methodological issues in demand models: self-selection and geographical scales. Hence an ordinal model is estimated as base of a joint model of residential location and trip-frequency. The next section describes the case study and data collection to achieve the objectives.

## DATA

### *Case study*

Madrid City has a population of 3.2 million inhabitants and 604 km<sup>2</sup> of area. Madrid Metropolitan Area (MMA) comprises the city of Madrid and forty surrounding municipalities. Madrid Metropolitan Area has a population of 6.7 million inhabitants and 1,935.97 km<sup>2</sup>. Both Madrid City and municipalities of the MMA are administratively divided in a number of districts. Madrid City is divided into 21 districts and 128 wards. Population density is quite different: Madrid City has a population density of 5,390 inhabitants/km<sup>2</sup>; while the MMA has a population density of 2,625 inhabitants/km<sup>2</sup>.



FIGURE 1 illustrates the case study, which is composed by 2 districts and 1 ward. The objective of choosing 3 different (TAZ) is to capture the *neighbourhood type effect*. In this context, this effect occurs when a specific mobility pattern is exhibited by citizens that live in the same neighbourhood. Hence, this is a suitable case study for analyzing urban sprawl due to new urban development and substantial changes in mobility patterns in the last years. The case study is called: *CBD, Urban and Suburban*, and described as follows:

- *CBD*: the study area corresponds to *Chamberí* ward, one of the 22 wards of the Central Business District of Madrid. It is a traditional neighbourhood where several historical buildings are located and where people live mainly in apartments. The area is characterised by good public transport service (bus, metro and rail) and by a gross income level that ranks the 4th of the 22 neighbourhoods of Madrid City.
- *Urban*: the study area corresponds to a district of *Pozuelo de Alarcón*, located 15 km west to the Madrid CBD but it is inside Madrid City. Public transport service is limited in this area. Urban residents tend to a car-oriented mobility pattern, living in single family houses or detached houses. The average income level of Pozuelo ranks the highest amongst the municipalities of the Region of Madrid.
- *Suburban*: the study area corresponds a district of *Algete*, located 30 km north-east to the Madrid CBD, in the Metropolitan Ring. Algete has lower available gross income and fewer transit services than the other two selected areas. Average income level of Algete ranks the 15th amongst the 179 municipalities of the Region of Madrid.

A household travel survey was conducted in 2006-2007 in the districts and ward mentioned before. The questionnaire consists in a trip diary, every individual older than four years old specifies the number of trips, purpose, travel time, transport mode, origin and destination of each trip carried out during the previous working day. Origin and destinations are specified with street and number. Only one individual per household fulfil the household characteristics. Data was collected with the aim of analysing the influence of the type of questionnaire on mobility patterns (34). Then, two questionnaires were used: activity based and travel based. The sample was balanced for each type of questionnaire. A total of 943 individuals were interviewed, from 345 households, distributed as follows: 288 from CBD, 372 from Urban and 283 from Suburban.

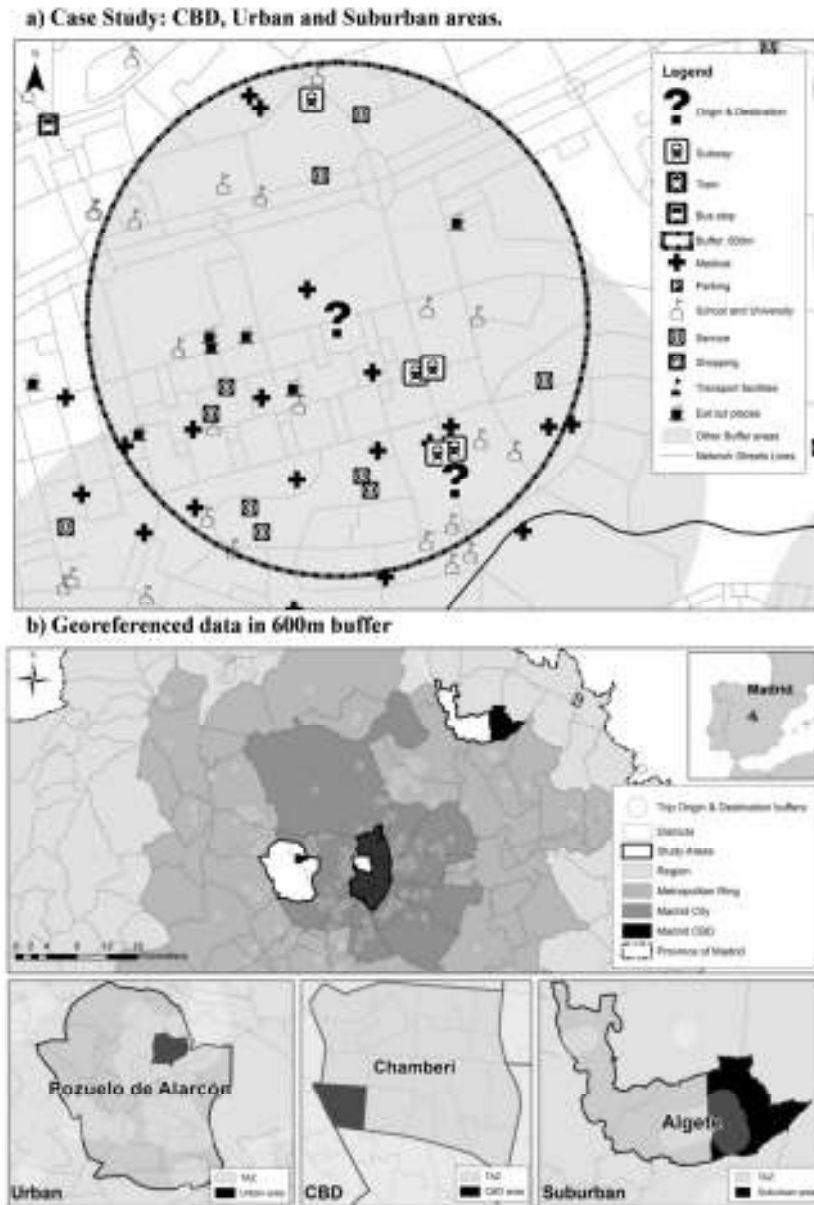


FIGURE 1 a) Case Study: CBD, Urban and Suburban areas. b) Georeferenced data in 600m buffer

### Data Extraction

BE variables were measured at three different zone levels: by municipalities, by District and in a radius of 600 meters around the household location (called “buffer”). Whenever it was possible, the same variable was measured at all the three levels, but in some cases the variables were available or made sense, only at some of the three levels. Variables at Municipality and District levels were computed using data from INE database. The residential level was defined as the area in the 600 meters radius around the residence of each person interviewed. This buffer was defined based on based on the distance that can be covered in about a ten-minute walk, and pedestrian accessibility studies conducted of the Public Transport Authority of Madrid (Gutiérrez and García-Palomares 2008).

Household locations of each person interviewed were address-matched with the INE database and integrated into a GIS. This data base includes both public transport supply and network street data. As FIGURE 2 shows on the left, households and trips (origins and destinations) were imported as comma delimited (csv) files tables and georeferenced using Google Earth Plus 5.1. and exported to ArcGIS 10 software. A similar process was carried out to locate facilities from Yellow Pages Directory. The Network Analyst of ArcGIS was used to obtain streets intersections from network street database. These intersections were characterized by number of crossing streets by splitting network street lines. Cul de sac intersections were calculated by feature vertices to points (using dangle command) tool from network street lines. Finally, in order to aggregate data inputs around origin and destinations, buffer features were intersected with land-use measures, and summarized by buffer identifier. . As result, we obtained *zonal activity opportunity* and *zonal transportation network measures*. FIGURE 1 illustrates an example of georeferenced origin/destination point in a buffer of 600m.

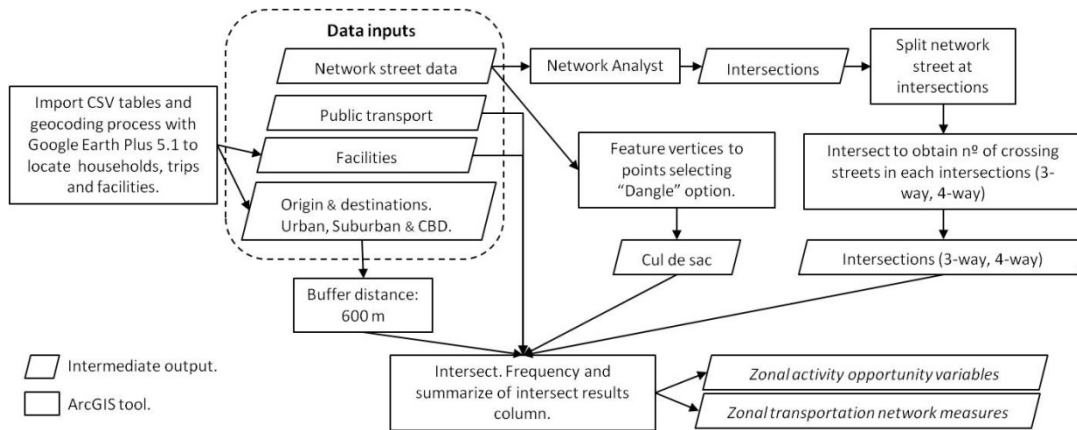


FIGURE 2. Sequence of GIS operations used to define Zonal activity opportunity variables and Zonal Land-use structure variables

## Measures

Data obtained from address-match was used to compute a set of BE measures for three geographical scales: municipality, district and buffer area, including:

1. *Zonal land-use structure variables (ZLUS)*: such as, fraction of residential, commercial and industrial land-use over urban land-use; and ratio of commercial, residential and industrial land-uses between origin and destination in hectares. These variables are only calculated by district and municipality.
2. *Zonal activity opportunity variables (ZAO)*: such as gross domestic product, percentage of workers at origin and destination, business establishment and facilities per square kilometre (service, eat-out places, medical, parking and schools). Particularly, the percentage of workers at origin and destination was calculated based on the total inhabitants for the origin/destination by municipality. And similarly, the ratio of worker is calculated dividing the percentage of workers at origin by the percentage of workers at destination. These variables were extracted from INE data base and GIS layers.

3. *Zonal transportation network measures (ZTN)*: this group includes variables such as intersection density (3-way, 4 way intersections and cul-de-sacs), number of bus stops, metro stations and rail stations per squared kilometres. The variables were extracted from GIS layers as explained previously.

## EMPIRICAL ANALYSIS

### Model Specification

After the data extraction, most appropriate indicators are used to estimate trip frequency. Given the ordinal nature of number of trips, the model estimated is an ordered logit model, which is a member of the ordinal model's family (Zavoina 1975). The ordinal logit model is obtained under the assumption that the error term  $\varepsilon$  is distributed logistic instead of standard normal. The dependent variable is unobserved ( $y_n^*$ ), but it is materialized in the number of trips collected from the trip diary. Let  $y_n^*$  be the utility that an individual  $n$  facing ordinal decision associates to alternative  $i$ . The ordinal logit probability model is based on a latent regression of the Alternative Specific Constant (ASC) and the explanatory variables:

$$y_{nqh}^* = ASC + \sum_{c=1,C} \lambda_c S_{cn} + \sum_{k=1,K} \beta_k Z_{kqh} + \varepsilon_n$$

**Eq. 1**

Where  $y_{nqh}^*$  is the unobserved dependent variable decomposed into the usual systematic and random components, as always,  $n$  denotes the individual.  $S_n$  a vector of SE characteristics,  $Z_q$  a vector of BE attributes measured at a spatial scale  $h$  for each origin or destination  $q$ . This vector can be written as  $Z_q = (U_q, M_q, F_q)$ , because it includes three types of attributes (or sub-vectors of attributes):

$U_q$  is a sub-vector of BE characteristics associated with the land-use variables (*ZLUS*)

$M_q$  is a sub-vector of BE characteristics associated with public transport at origin and destination (*ZTN*).

$F_q$  is a sub-vector of BE characteristics associated with the commercial urban retails and facilities (*ZAO*).

Finally,  $\varepsilon_n$  represents the error term logistic distributed with zero mean and standard deviation  $\sigma_\varepsilon$ . All models were estimated with the software package BIOGEME (37).

Since an ordered logit model is estimated, the thresholds are defined as follows:

$$\begin{aligned} y_n = 0 & \quad \text{if } y_{nqh}^* < \mu_0, \text{ then Trip} = 0 \\ y_n = 1 & \quad \text{if } \mu_0 < y_{nqh}^* \leq \mu_1, \text{ then } 1 \leq \text{Trip} \leq 2 \\ y_n = 3 & \quad \text{if } \mu_1 < y_{nqh}^* \leq \mu_2, \text{ then } 3 \leq \text{Trip} \leq 4 \\ y_n = 4 & \quad \text{if } y_{nqh}^* > \mu_2, \text{ then Trip} > 4 \end{aligned}$$

**Eq. 2**

Where  $\mu$ 's are threshold parameters estimated together with the vectors of parameters  $\beta$  and  $\lambda$ .  $y_{nqh}^*$  is decomposed into observed ( $\beta_k Z_{kqh}$  and  $\lambda_c S_{cn}$ ) and unobserved components ( $\varepsilon_n$ ). Then, the probability of observing  $y_n$  taking the value  $j$  is given by:

$$\begin{aligned} P_{nj} &= Prob(y_{nqh}^* < \mu_i) \\ &= Prob((\lambda_c S_{cn} + \beta_k Z_{kqh}) < \mu_i) \\ &= Prob(\varepsilon_n < \mu_i - (\lambda_c S_{cn} + \beta_k Z_{kqh})) = \frac{e^{\mu_j - (\lambda_c S_{cn} + \beta_k Z_{kqh})}}{1 + e^{\mu_j - (\lambda_c S_{cn} + \beta_k Z_{kqh})}} \end{aligned} \quad \text{Eq. 3}$$

Similarly, the probability of choosing the alternative  $j=J-1$  can be explained as:

$$P_{nj} = \frac{e^{\mu_{J-1} - (\lambda_c S_{cn} + \beta_k Z_{kqh})}}{1 + e^{\mu_{J-1} - (\lambda_c S_{cn} + \beta_k Z_{kqh})}} - \frac{e^{\mu_J - (\lambda_c S_{cn} + \beta_k Z_{kqh})}}{1 + e^{\mu_J - (\lambda_c S_{cn} + \beta_k Z_{kqh})}} \quad \text{Eq. 4}$$

The ordinal logit structure can be optimized via maximum likelihood using procedures for standard logit estimation but treated as binary logit. This model represents the choice as the outcome of a sequence of binary decisions. Thus, it is not based on global utility maximization of values, one alternative encompasses several values. Then, the decision consists on whether "take one more" or accept the current value, instead of choosing among alternatives. The maximization stops once the first local maximum is reached.

### Self-Selection Test

A joint mixed ordered model is used to estimate simultaneously the Residential location Choice ( $RC_n$ ) and Number of Trips ( $y_{nq}^*$ ). Then, a joint probability bundle is chosen and modelled. This certainly reflects a sequential structure on multiple decisions, different to other structures, such as nested logit.

As described in Bath and Guo (2007), self-selection can be accounted for through the effect of SE characteristics and the effect of unobserved heterogeneity common to both Number of Trips ( $y_{nq}^*$ ) and the Residential location Choice ( $RC_{nq}$ ) models. The sub-index  $h$  is removed in this set of equations because not every component of BE is measured at three spatial scales.

The equations can then be rewritten as:

$$\begin{aligned} y_{nq}^* &= \sum_{c=1,C} \lambda_c S_{cn} + \sum_{k=1,K} \gamma_k Z_{kq} + \sum_{l=1,L} \psi_{ln} Z_{lq} + \sum_{j=1,J} \omega_{jn} Z_{jq} + \varepsilon_n \\ RC_{nq} &= \sum_{c=1,C} \lambda_c S_{cn} + \sum_{k=1,K} \delta_k Z_{kq} + \sum_{d=1,D} \varphi_{dn} Z_{dn} + \sum_{j=1,J} \omega_{jn} Z_{jq} + \xi_n \end{aligned} \quad \text{Eq. 5}$$

Where  $\gamma_k$  and  $\delta_k$  are vectors of fixed parameters for BE attributes  $U_q$ ;  $\psi_{ln}$  and  $\varphi_{dn}$  are vectors of specific estimated parameters on trip frequency or residential location choice respectively, with size  $l$  and  $d$ , to be estimated for a sub-set of  $F_q$  and  $M_q$  vectors of attributes; and  $\omega_{jn}$  is a common vector on trip frequency and residential location choice, estimated for a sub set of or  $F_q$  and  $M_q$  vectors of attributes.

## RESULTS

After an exhaustive exploratory analysis of BE attributes, four model specifications were estimated in order to find the optimum model. During the estimation, in all cases, SE variables became irrelevant after adding BE variables. An important finding from here is the relative importance of BE attributes in respect to SE characteristics in travel demand models. Particularly, BE attributes contribute to more precise representation of trip frequency. TABLE 1 shows the model results for model specifications 1 to 4. Both parameters and t-test are reported in the table.

. We firstly focus the attention on the level of significance of 10 parameters, which are: 5 parameters of ZTN; 10 parameters of ZAO, which includes facilities and street network parameters. These parameters are included in all models and all geographical scales. The type of intersection could be included only in the model estimated at buffer level, but it was never significant. Specification 4 shows the highest number relevant variables, 6 relevant parameters at 95% level of confidence. However, Specification 3 shows a no significant constant, which indicates that the attributes included in the specification represent better the average phenomenon. Contrary, the model in case 4 shows a highly significant constant, it indicates a lot of disturbance, which is still lacking of explanation.

From a model fit standpoint, BE attributes calculated at buffer scale produce superior models. Rho square adjusted is higher in the buffer model (0.406), than in the municipality (0.404) and district model (0.227). Among models estimated with Specification 3, buffer scale presents the largest t-test values. It may indicate that statistical models based on walkable scales are more precise. This result is consistent with a number of studies on spatial dimensions of neighbourhood effects, see for example Spielman and Yoo (2009) and Krizek. (2003). We now focus on the results from model specification 3. In the comparison among the three spatial levels, it seems that the ratio of percentage of land-use between origin and destination is measured much better at district level than at municipality and residential level. Public transport measures are also more significant when measured at municipality level and buffer level, than district one. It must also be noted that some of these at the district level have an opposite sign than at municipality and buffer level. According to the results of models estimated the frequency is elastic to changes in accessibility but models at Municipality and Buffer scale correctly simulate that the demand increases, while results of District scale model indicates that travel demand decreases. This result is of course worrying because it implies an opposite forecast. The model at district level is then not correct.

ZAO, specifically number of schools, service, medical-oriented and parking facilities at origin of the trip tend to increase the number of trips. This result is consistent among three spatial levels. One of the most important indicators of zonal facilities is the number of schools within the buffer area. Surprisingly, all the BE measures related with the intersection characteristics, and measured at residential level, are relevant and negative. 5-way intersections represent dispersed street network, as consequence lower number of trips.

Finally, the estimated threshold are all significant at 95% confidence level (t-test higher than 1.96) in all the estimated models. These are also in ascending order showing that the intended order is correct (Andrich, de Jong et al. 1997). These thresholds in fact represent the demarcation points on the continuous latent propensity scale that identify the observed discrete values of person-trips.

### *Test for Self-selection*

Model results are used to verify the effect of preferences in residential location choice and trip frequency, via self-selection test.

The test for self-selection was carried out through estimating a joint mixed ordered model. The model specification is based on Specification 3. TABLE 2 shows the model results of the joint ordered model. Land-use variables at municipality level were assumed to be fixed parameters because of the lack of variability of these measures. Facilities and public transport parameters were firstly tested as specific random parameters for  $y_{nq}^*$  and  $RC_n^*$ . We choose the appropriate specification based on the coefficient: equal values of estimated parameters indicate shared unobserved factors, and then common random parameters. According to this, statistically significant standard deviation of common random parameters indicate the association between  $y_{nq}^*$  and  $RC_n^*$ , and as result the self-selection.

Analysing the preference for BE attributes common between residential location choices and number of trips, in general, the results show that testing for sensitivity in BE attributes is important for both decisions processes. Results for self-selection effect can be accounted for in the statistically significant  $\sigma_\omega^2$  estimates.

It can be observed in the municipality and district models, that number of schools and its error term are statistically significant. This indicates that a shared sensitivity is accounted for individuals during the joint decision process. Thus, individuals perceive the number of schools around their origin and destination area as attractive for number of trips and also perceive it as attractive factor for residential choice. Reasonably, the sign is positive, indicating that people who perceive the number of schools as attractive for undertaking more trips are more likely to choose CBD as residential location. This indicates that self-selection exists in this sample and must be taken into account to estimate trip frequency. Hence, models in TABLE 2 are superior to models in TABLE 1.

The buffer scale shows more satisfactory and stable results. Hence, the buffer scale is the most efficient scale to conduct an analysis of self-selection in this context. Again, at this scale, results show that households are more likely to locate in zones with better schooling and transit opportunities. As stated in the introduction, geographical scale proved to be important on travel demand estimation, since we obtain different results.

Residential self-selection and geographical scales in trip frequency

LA PAIX, Lissy; MONZÓN, Andrés; CHERCHI, Elisabetta and LORO, Manuel

**TABLE 1 Results for Model Specifications 1 to 4**

	Model Specification 1			Model Specification 2			Model Specification 3			Model Specification 4		
Spatial Scale	Municipality	District	Buffer	Municipality	District	Buffer	Municipality	District	Buffer	Municipality	District	Buffer
Variables	Value (t-test)*	Value (t-test)*	Value (t-test)*	Value (t-test)*	Value (t-test)*	Value (t-test)*	Value (t-test)*	Value (t-test)*	Value (t-test)*	Value (t-test)*	Value (t-test)*	Value (t-test)*
<b>ASC2</b>	1.620 <b>(4.74)</b>	-0.213 <b>(-0.55)</b>	-0.673 <b>(-1.62)</b>	-0.444 <b>(-0.62)</b>	-0.645 <b>(-1.06)</b>	0.764 <b>(2.19)</b>	-0.026 <b>(-0.04)</b>	0.481 <b>(0.76)</b>	0.459 <b>(1.40)</b>	6.830 <b>(2.87)</b>	16.900 <b>(7.45)</b>	<b>3.920</b> <b>(5.49)</b>
<i>Socio-economic variables</i>												
<b>Workers</b>	0.389 <b>(2.5)</b>	-0.153 <b>(-0.88)</b>	0.574 <b>(3.62)</b>	0.113 <b>(0.68)</b>	0.234 <b>(1.54)</b>	0.526 <b>(3.29)</b>	0.114 <b>(0.68)</b>	0.338 <b>(2.26)</b>	0.545 <b>(3.42)</b>	0.106 <b>(0.64)</b>	0.213 <b>(1.41)</b>	0.004 <b>(0.02)</b>
<b>Driver</b>	0.400 <b>(2.23)</b>	0.558 <b>(2.72)</b>	0.526 <b>(3.19)</b>	0.578 <b>(2.99)</b>	0.344 <b>(2.05)</b>	0.616 <b>(3.69)</b>	0.543 <b>(2.78)</b>	0.320 <b>(1.91)</b>	0.636 <b>(3.82)</b>	0.484 <b>(2.54)</b>	0.344 <b>(2.04)</b>	0.597 <b>(2.82)</b>
<i>Zonal land-use structure variables (ZLUS)</i>												
<b>Ratio OD Commercial</b>	0.066 <b>(0.99)</b>	0.127 <b>(2.37)</b>	0.174 <b>(3.24)</b>									
<b>Ratio OD Industrial</b>	-0.059 <b>(-2.39)</b>	-0.073 <b>(-2.75)</b>	-0.052 <b>(-1.93)</b>									
<b>Ratio OD Residential</b>	0.185 <b>(0.84)</b>	-1.320 <b>(-5.43)</b>	0.251 <b>(-1.24)</b>									
<b>Residential at origin</b>				-0.222 <b>(-0.28)</b>	-2.620 <b>(-4.05)</b>	-1.470 <b>(-2.12)</b>						
<b>Commercial origin</b>							9.140 <b>(1.77)</b>	13.700 <b>(2.45)</b>				
<b>Commercial destination</b>							-0.806 <b>(-0.100)</b>		-17.200 <b>(-2.43)</b>	-21.500 <b>(-3.01)</b>		-25.700 <b>(-3.34)</b>
<b>Industrial at origin</b>				6.990 <b>(4.70)</b>	-0.675 <b>(-0.51)</b>	-0.236 <b>(-0.17)</b>	7.550 <b>(4.96)</b>	0.521 <b>(0.37)</b>			24.600 <b>(4.37)</b>	40.600 <b>(7.64)</b>
<b>Industrial at destination</b>				4.860 <b>(3.64)</b>	-4.610 <b>(-3.78)</b>	0.151 <b>(0.14)</b>	4.830 <b>(4.15)</b>	-3.170 <b>(-2.9)</b>	1.070 <b>(1.16)</b>	4.500 <b>(3.82)</b>	-2.980 <b>(-2.74)</b>	3.690 <b>(3.39)</b>
<i>Zonal activity opportunity variables (ZAO):</i>												
<b>Workers/sq km</b>	3.400 <b>(2.09)</b>	4.190 <b>(2.41)</b>	0.601 <b>(0.14)</b>									
<b>% Workers at</b>				-0.896 <b>(-0.42)</b>	6.880 <b>(3.86)</b>	-0.308 <b>(-0.20)</b>	-0.951 <b>(-0.50)</b>	3.850 <b>(2.36)</b>	-2.010 <b>(-1.68)</b>	-4.360 <b>(-2.5)</b>	1.640 <b>(1.12)</b>	-10.500 <b>(-8.39)</b>

13<sup>th</sup> WCTR, July 15-18, 2010 – Rio, Brazil



Residential self-selection and geographical scales in trip frequency

LA PAIX, Lissy; MONZÓN, Andrés; CHERCHI, Elisabetta and LORO, Manuel

	Model Specification 1			Model Specification 2			Model Specification 3			Model Specification 4		
Spatial Scale	Municipality	District	Buffer	Municipality	District	Buffer	Municipality	District	Buffer	Municipality	District	Buffer
Destination												
Ratio GDP per capita O/D	0.522 <b>(2.51)</b>	0.231 (1.08)	-0.066 (-0.250)									
GDP per capita at origin										1.390E-04 <b>(3.05)</b>	3.350E-04 <b>(7.70)</b>	-2.190E-05 <b>(-1.76)</b>
GDP per capita at destination				-5.89E-05 <b>(-5.9)</b>	-1.31E-05 (-1.51)	-9.8 E-06 (-1.10)	-5.83E-05 <b>(-5.4)</b>	-9.00E-06 -1.04	3.45E-06 0.33			
Medical	0.324 <b>(3.16)</b>	-0.001 (-0.10)	0.142 (0.92)	-0.015 (-0.15)	0.128 <b>(1.99)</b>	-0.027 (-0.20)			-0.053 (-0.39)	0.296 <b>(1.96)</b>	0.325 <b>(4.54)</b>	0.046 (0.29)
School			0.181 <b>(6.13)</b>	0.235 <b>(3.53)</b>	0.438 <b>(4.11)</b>	0.189 <b>(5.85)</b>	0.251 <b>(3.6)</b>	0.327 <b>(3.66)</b>	0.187 <b>(6.02)</b>	0.231 <b>(3.91)</b>	0.325 <b>(4.02)</b>	0.082 <b>(2.74)</b>
Service		0.022 (0.72)	0.006 (0.29)		0.014 (0.52)	-0.005 (-0.24)		0.108 <b>(2.92)</b>	-0.007 (-0.37)		0.136 <b>(3.94)</b>	-0.035 (-1.35)
Parking		0.157 (1.31)	0.571 <b>(2.11)</b>		0.154 (1.44)	0.291 (1.13)			0.313 (1.23)		0.520 <b>(4.44)</b>	0.431 (1.63)
Eat out place	0.001 (0.07)		0.175 <b>(3.53)</b>	-0.011 (-0.77)		0.110 <b>(2.20)</b>	-0.006 (-0.43)		0.102 <b>(1.97)</b>	0.014 (1.030)		0.096 (1.88)
<i>Zonal transportation network measures (ZTN)</i>												
Interurban bus lines			14.800 <b>(18.07)</b>									12.800 <b>(12.23)</b>
Bus stops	0.155 <b>(10.66)</b>	-0.074 <b>(-7.05)</b>	0.154 <b>(11.17)</b>	0.210 <b>(11.86)</b>	-0.030 <b>(-2.98)</b>	0.157 <b>(11.08)</b>	0.201 <b>(9.3)</b>	-0.037 <b>(-3.59)</b>	0.166 <b>(10.92)</b>	0.207 <b>(10.02)</b>	-0.029 <b>(-3.49)</b>	0.076 <b>(4.31)</b>
Metro stations	3.990 <b>(7.13)</b>	-0.083 <b>(-5.02)</b>	0.125 <b>(7.22)</b>	6.100 <b>(9.94)</b>	-0.102 <b>(-6.36)</b>	0.138 <b>(7.8)</b>	6.270 <b>(8.44)</b>	-0.055 <b>(-3.24)</b>	0.130 <b>(7.19)</b>	4.850 <b>(6.81)</b>		0.027 (1.25)
Rail stations		0.634 <b>(3.3)</b>	-0.275 (-1.58)		0.735 <b>(4.36)</b>	-0.146 (-0.84)		0.863 <b>(5.26)</b>	-0.155 (-0.90)		0.635 <b>(4.13)</b>	-0.522 <b>(-2.51)</b>
4street			-0.866 (-1.43)									-1.360 <b>(-2.38)</b>

Residential self-selection and geographical scales in trip frequency

LA PAIX, Lissy; MONZÓN, Andrés; CHERCHI, Elisabetta and LORO, Manuel

	Model Specification 1			Model Specification 2			Model Specification 3			Model Specification 4		
Spatial Scale	Municipality	District	Buffer	Municipality	District	Buffer	Municipality	District	Buffer	Municipality	District	Buffer
5street			-6.750 (-1.85)			-4.470 (-1.11)			-3.880 (-1.06)			
<i>Thresholds</i>												
tau2	3.630 <b>(22.14)</b>	4.510 <b>(20.15)</b>	3.460 <b>(22.54)</b>	3.910 <b>(22.14)</b>	2.810 <b>(24.48)</b>	3.380 <b>(22.63)</b>	3.910 <b>(22.19)</b>	2.740 <b>(24.82)</b>	3.390 <b>(22.49)</b>	3.880 <b>(22.16)</b>	2.820 <b>(24.68)</b>	4.530 <b>(24.06)</b>
tau3	5.560 <b>(25.2)</b>	6.420 <b>(24)</b>	1.890 <b>(12.47)</b>	5.820 <b>(25.45)</b>	4.590 <b>(25.64)</b>	1.870 <b>(12.55)</b>	5.810 <b>(25.48)</b>	4.520 <b>(25.67)</b>	1.880 <b>(12.6)</b>	5.790 <b>(25.42)</b>	4.610 <b>(25.77)</b>	1.900 <b>(12.49)</b>
<i>Fit measures</i>												
Number of estimated parameters	14	17	20	15	17	19	15	15	18	15	15	19
Adjusted rho Square	0.347	0.449	0.412	0.403	0.236	0.403	0.404	0.227	0.406	0.393	0.242	0.526

\*Significant t-test, higher than 1.96, reported in parenthesis and bold letters below the parameter.

**TABLE 2 Results for joint model of self-selection test**

	Municipality		District		Buffer	
	Value	Robust t-test	Value	Robust t-test	Value	Robust t-test
b_meanAtt	-0.04	-0.09	-0.42	-1.04	-0.20	-0.66
<i>Socio-demographic variables</i>			-0.57	-2.52	-0.41	-1.89
Driver	0.45	2.81	0.37	2.19	0.38	2.68
Worker	0.16	1.13	0.29	1.72	0.36	2.58
Threshold value 1 (between category 1 and 2)	3.96	24.28	3.55	17.73	3.19	23.74
Increment of threshold value 1 (between category 2 and 3)	1.94	12.07	2.20	10.50	1.91	11.77

13<sup>th</sup> WCTR, July 15-18, 2010 – Rio, Brazil

Residential self-selection and geographical scales in trip frequency

LA PAIX, Lissy; MONZÓN, Andrés; CHERCHI, Elisabetta and LORO, Manuel

	Municipality		District		Buffer	
	Value	Robust t-test	Value	Robust t-test	Value	Robust t-test
<b>Random Parameters (Y*)</b>						
Number of eat-out places per square kilometers	-0.03	-2.35	0.05	2.07	0.18	3.49
Standard Deviation	0.00	-0.53	0.03	0.39	0.07	0.89
<b>Common Random BE Parameters (U* Y*)</b>						
Number of bus stops per squared kilometer at destination	0.12	8.33	0.06	6.46	0.09	8.31
Standard Deviation	2.49E-04	0.46	7.04E-04	0.56	0.05	4.79
Number of metro stations per squared kilometer at destination	5.05	15.13	0.09	6.62	0.14	10.36
Standard Deviation	-2.76E-03	-0.15	-1.04E-04	-0.28	-3.56E-03	-0.87
Number of schools and university per square kilometres	0.81	4.11	2.06	6.47	0.18	5.12
Standard Deviation	-0.38	-3.88	-1.20	-6.39	-0.06	-1.96
<b>Fixed Parameters Y* of BE</b>						
Percent of Industrial Land-use at origin	3.95	3.64	3.09	2.09	2.11	2.69
Percent of Industrial Land-use at origin	11.10	3.03	22.50	4.67	16.30	4.35
Gross Domestic Product (GDP) at municipality of destination	-4.53E-03	-5.90	-4.44E-03	-4.15	-1.76E-03	-2.91
<b>ASC-CBD</b>						
ASC-CBD	0.43	1.52	-0.04	-0.11	1.12	4.87
<b>ASC-Urban</b>						
ASC-Urban	-1.06	-3.97	-1.70	-5.31	-0.50	-2.02
<b>Socio-demographic variables</b>						
Female-married CBD	-0.52	-2.05	-0.57	-2.70	-0.62	-2.83
Female-married Urban	-0.32	-1.43	-0.38	-1.95	-0.45	-2.22
Holding driver license. Yes=1; otherwise 1 (Specific for CBD choice)	-0.22	-1.02	-0.03	-0.15		
Number of car per household, including company car (from 0 to 4) (Specific for CBD choice)	-0.81	-6.84	-0.82	-7.44	-0.84	-7.97

13<sup>th</sup> WCTR, July 15-18, 2010 – Rio, Brazil

Residential self-selection and geographical scales in trip frequency

LA PAIX, Lissy; MONZÓN, Andrés; CHERCHI, Elisabetta and LORO, Manuel

	Municipality		District		Buffer	
	Value	Robust t-test	Value	Robust t-test	Value	Robust t-test
Number of car per household, including company car (from 0 to 4) (Specific for Urban choice)	0.27	2.91	0.27	3.00	0.25	2.94
Occupation is worker, Yes=1; otherwise 0 (Specific for Urban choice)	-0.26	-1.56	-0.14	-0.90	-0.07	-0.48
Travel time for the longest working Trip	-0.05	-7.44	-0.01	-3.54	-0.02	-5.09
<b>Fixed Parameters (U) of BE</b>						
Number of parking facilities per square kilometres	0.02	3.48				

**TABLE 3 Self-selection Models for District and Municipality Geographical Scales**

## CONCLUSIONS

This paper analyzes the effect of different spatial scales and the detail of indicators for representing BE. The methodology developed here contributes to the better understanding of the relationship between self-selection and the scale to represent BE attributes. The analysis shows that both geographical scale and self-selection are crucial on the analysis of the relationship BE and TB.

Smaller geographical scales are more suitable to represent the individual sensitivity to BE attributes than large administrative units. Then, a “buffer” scale represent what people really matter in their trip decisions. The same model specification produces different effects at different spatial scales; particularly for transit measures. Hence, transportation planners must design tools and transport measures according to geographical scales.

Through the selected 3 neighbourhood types, the model shows that neighbourhood effect is relevant for analyzing TB. Thus, the residence zones with higher density of opportunities/activities produce more trips.

The estimation of standard ordinal models of trip frequency shows that the best result from a combination of geographical scales. As can be seen, it is possible to combine geographical scales by including municipality measures in the model estimation. Consistent with Krizek (2003) residential location decisions may be influenced by the character of the particular neighbourhood and the position of the neighbourhood within its region. This means that both buffer and larger scales are important up to some level of analysis.

Regarding to the test for self-selection effects, it indicates that individuals' preferences are associated to trips decisions. Therefore self-selection should be considered in travel demand models. The analytical framework is an explicit representation of the simultaneous decision process which includes residential location choice and number of trips. It has proved to be more robust than trip frequency ordinal models.

As future research, tour based analysis can be applied instead of trip frequency. Trip-related variables were calculated for the first trip during a full-day tour, an analysis of these variables by trip must be interesting as well.

## ACKNOWLEDGMENTS

The authors would like to thank to the Prof. Javier Gutiérrez Puebla at the Complutense University of Madrid, for providing the transportation network in Arc GIS.

## REFERENCES

- Abreu e Silva, J., T. F. Golob, et al. (1977). "The effects of land use characteristics on residence and employment location and travel behaviour of urban adult workers." Transportation Research Record(Journal Article): 121-131.
- Agyemang-Duah, K., W. P. Anderson, et al. (1995). "Trip generation for shopping travel." Transportation Research Record(1493): 12-20.

Residential self-selection and geographical scales in trip frequency

LA PAIX, Lissy; MONZÓN, Andrés; CHERCHI, Elisabetta and LORO, Manuel

- Andrich, D., J. de Jong, et al. (1997). Diagnostic opportunities with the Rasch model for ordered response categories. Applications of latent trait and latent class models in the social sciences. J. Rost and R. Langeheine. Münster, Germany. **Waxman Verlag**: pp. 58-68.
- Berry, S. (1995). "Automobile prices in market equilibrium." Econometrica **63**(4): 841.
- Bhat, C. (1999). "An analysis of evening commute stop-making behavior using repeated choice observations from a multi-day survey." Transportation Research Part B **33**(7): 495-510.
- Bhat, C. and J. Y. Guo (2007). "A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels." Transportation Research Part B: Methodological **41**(5): 506-526.
- Bhat, C. and S. Srinivasan (2005). "A multidimensional mixed ordered-response model for analyzing weekend activity participation." Transportation Research Part B: Methodological **39**(3): 255-278.
- Boarnet, M. and R. Crane (2001). "The influence of land use on travel behavior: specification and estimation strategies." Transportation Research Part A: Policy and Practice **35**(9): 823-845.
- Boarnet, M. and R. Crane (2001). "The influence of land use on travel behavior: specification and estimation strategies." Transportation Research Part A **35**(9): 823-845.
- Boarnet, M. and S. Sarmiento (1998). "Can land-use policy really affect travel behaviour? A study of the link between non-work travel and land-use characteristics." Urban Studies **35**(7): 1155-1169.
- Bohte, W. (2010). Residential self-selection and travel: The relationship between travel-related attitudes, built environment characteristics and travel behaviour, TU Delft.
- Brownstone, D. (2008). "Key Relationships Between the Built Environment and VMT." Transportation Research Board(Journal Article): 2-4.
- Cao, X., P. L. Mokhtarian, et al. (2009). "Examining the impacts of residential self-selection on travel behaviour: a focus on empirical findings." Transport Reviews **29**(3): 359-395.
- Cao, X., P. L. Mokhtarian, et al. (2009). "Examining the Impacts of Residential Self-Selection on Travel Behaviour: A Focus on Empirical Findings." Transport Reviews: A Transnational Transdisciplinary Journal **29**(3): 359 - 395.
- Cervero, R. and R. Gorham (1995). "Commuting in transit versus automobile neighborhoods." (Journal Article).

Residential self-selection and geographical scales in trip frequency

LA PAIX, Lissy; MONZÓN, Andrés; CHERCHI, Elisabetta and LORO, Manuel

- Coulton, C. J., J. Korbin, et al. (2001). "Mapping residents' perceptions of neighborhood boundaries: a methodological note." American Journal of Community Psychology **29**(2): 371.
- Dieleman, F. M., M. Dijst, et al. (2002). "Urban form and travel behaviour: micro-level household attributes and residential context." Urban Studies **39**(3): 507.
- Ecotec (1993). Reducing transport emissions through land use planning. London, HMSO.
- Ewing, R., M. B. DeAnna, et al. (1996). "Land use impacts on trip generation rates." Transportation Research Record: Journal of the Transportation Research Board **1518**(-1): 1-6.
- Ewing, R., R. Pendall, et al. (2003). "Measuring sprawl and its transportation impacts." Transportation Research Record: Journal of the Transportation Research Board **1831**(-1): 175-183.
- Farber, S. and A. Paez (2009). "My car, my friends, and me: a preliminary analysis of automobility and social activity participation." Journal of Transport Geography **17**(Journal Article): 216-225.
- Frank, L., M. Bradley, et al. (2008). "Urban form, travel time, and cost relationships with tour complexity and mode choice." Transportation **35**(1): 37-54.
- Frank, L. D. and G. Pivo (1994). "Impacts of mixed use and density on utilization of three modes of travel: single-occupant vehicle, transit, and walking." Transportation Research Record: Journal of the Transportation Research Board(1466): 44-52.
- Friedman, B., S. P. Gordon, et al. (1994). "Effect of neotraditional neighborhood design on travel characteristics." In Transportation Research Record: Journal of the Transportation Research Board, No. 1446, Transportation Research Board of the National Academies, Washington, D.C. **1466**(Journal Article): 63–70.
- García-Palomares, J. C. (2009). "Urban sprawl and travel to work: the case of the metropolitan area of Madrid." Journal of Transport Geography(Journal Article).
- Geurs, K. T. and B. Van Wee (2004). "Accessibility evaluation of land-use and transport strategies: review and research directions." Journal of Transport geography **12**: 127-140.
- Golledge, R. G. and T. Gärling (2003). Cognitive maps and urban travel. University of California Transportation Center.
- Guo, J. Y. and C. R. Bhat (2007). "Operationalizing the concept of neighborhood: Application to residential location choice analysis." Journal of Transport Geography **15**(1): 31-45.
- Gutiérrez, J. and J. C. García-Palomares (2008). "Distance-measure impacts on the calculation of transport service areas using GIS." Environment and Planning B: Planning and Design **35**(3): 480-503.

Residential self-selection and geographical scales in trip frequency

*LA PAIX, Lissy; MONZÓN, Andrés; CHERCHI, Elisabetta and LORO, Manuel*

- Handy, S., X. Cao, et al. (2005). "Correlation or causality between the built environment and travel behavior? Evidence from Northern California." Transportation Research Part D **10**(6): 427-444.
- Hanson, S. (1982). "The determinants of daily travel-activity patterns: relative location and sociodemographic factors." Urban Geography **3**(3): 179-202.
- Khattak, A. J. and D. Rodriguez (2005). "Travel behavior in neo-traditional neighborhood developments: A case study in USA." Transportation Research Part A: Policy and Practice **39**(6): 481-500.
- Kitamura, R., P. L. Mokhtarian, et al. (1997). "A micro-analysis of land use and travel in five neighborhoods in the San Francisco Bay Area." Transportation **24**(2): 125-158.
- Krizek, K. J. (2003). "Neighborhood services, trip purpose, and tour-based travel." Transportation **30**(4): 387-410.
- Krizek, K. J. (2003). "Operationalizing Neighborhood Accessibility for Land Use-Travel Behavior Research and Regional Modeling." Journal of Planning Education and Research **22**(3): 270-287.
- Krizek, K. J. (2003). "Residential relocation and changes in urban travel: does neighborhood-scale urban form matter." Journal of the American Planning Association **69**(3): 265.
- La Paix, L. (2012). Modelling the Impact of Built Environment, Geographical Scales and Latent Constructs on Individual Travel Behaviour. Doctoral Thesis. Dissertation/Thesis, Universidad Politécnica de Madrid.
- La Paix, L., A. Monzón, et al. Modelling the relationship between urban environment and travel behaviour: policy and indicators. MOBIL.TUM 2012 International Scientific Conference on Mobility and Transport.
- Louviere, J. (2005). "Recent progress on endogeneity in choice modeling." Marketing Letters **16**(3): 255.
- Mokhtarian, P. L. and X. Cao (2008). "Examining the impacts of residential self-selection on travel behavior: A focus on methodologies." Transportation Research Part B: Methodological **42**(3): 204-228.
- Naess, P. (2009). "Residential Self-Selection and Appropriate Control Variables in Land Use: Travel Studies." Transport Reviews: A Transnational Transdisciplinary Journal **29**(3): 293 - 324.
- Paez, A., D. Scott, et al. (2007). "Elderly mobility: Demographic and spatial analysis of trip making in the Hamilton CMA, Canada." Urban Studies **44**(1): 123-146.



Residential self-selection and geographical scales in trip frequency

*LA PAIX, Lissy; MONZÓN, Andrés; CHERCHI, Elisabetta and LORO, Manuel*

- Pinjari, A., R. Pendyala, et al. (2007). "Modeling residential sorting effects to understand the impact of the built environment on commute mode choice." Transportation **34**(5): 557-573.
- Pinjari, A. R., C. R. Bhat, et al. (2009). "Residential self-selection effects in an activity time-use behavior model." Transportation Research Part B: Methodological **43**(7): 729-748.
- Pinjari, A. R., R. M. Pendyala, et al. (2007). "Modeling residential sorting effects to understand the impact of the built environment on commute mode choice." Transportation **34**(5): 557-573.
- Salon (2009). "Neighborhoods, cars, and commuting in New York City: A discrete choice approach." Transportation Research Part E : Logistics and Transportation Review **43**(2): 180.
- Salon, D. (2009). Cars and the city: an investigation of transportation and residential location choices in New York City. Dissertation/Thesis, University of California Davis.
- Spielman, S. E. and E.-h. Yoo (2009). "The spatial dimensions of neighborhood effects." Social science & medicine **68**(6): 1098-1105.
- Stead, D. and S. Marshall (2001). "The relationships between urban form and travel patterns: an international review and evaluation." European Journal of Transport and Infrastructure Research **1**(2): 113.
- Van Acker, V. (2007). "The effects of the land use system on travel behavior: a structural equation modeling approach." Transportation Planning and Technology **30**(4): 331.
- Zavoina (1975). "A statistical model for the analysis of ordinal level dependent variables." The Journal of mathematical sociology **4**(1): 103.