



Modelling the spatial interactions between workplace and residential location



Ángel Ibeas^a, Ruben Cordera^c, Luigi dell'Olio^{a,*}, Pierluigi Coppola^b

^a Universidad de Cantabria, Av. de Los Castros s/n, 39005 Santander, Cantabria, Spain

^b Department of Enterprise Engineering, "Tor Vergata" University of Rome, Italy

^c ETRA Norte S.A, Santander, Spain

ARTICLE INFO

Article history:

Received 26 November 2011

Received in revised form 30 December 2012

Accepted 3 January 2013

Keywords:

Discrete choice models

Land-use–transport–interaction (LUTI)

Accessibility

ABSTRACT

The use of Multinomial Logit (MNL) models specification for the simulation of residential location have been often criticised due to the Independence of Irrelevant Alternatives hypothesis (IIA) which does not allow for the existence of spatial correlation between residential zones. Moreover, it is not clear when and to what extent the influence of the workplace zone and accessibility to employment affect the residential location choices made by households; in other word, whether the residing choice is conditional to the workplace, or vice versa; or if such choices (residence and work place) are joint.

In this paper, Nested Logit (NL) and Cross-Nested Logit models of residential location choice are specified and compared to MNL, to investigate the existence of spatial correlation between different locations. Furthermore, different assumptions are tested, considering the choice of residential zone and the joint choice of residential zone and work place zone.

The models were estimated for the urban area of Santander (Spain). The results indicate that the inclusion in the model specification, of the spatial correlation between zones fit the data significantly better. Home-work journey times were a statistically significant factor in household location choice, whereas accessibility to employment had the correct sign but it was not statistically significant.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction and objectives

Land use and transport interaction models (LUTI models) are based on the hypothesis of a strong interrelationship between the population location pattern, the economic activities location pattern and the functioning of the transport system (Barra, 1989). Classic LUTI models such as developed by Lowry (1964) postulate that home-work journey costs are one of the basic factors in explaining and simulating household location. Other more recent models (Waddell, 2002) also consider measurements of accessibility to employment in determining how attractive households find specific locations. However, only a limited number of case studies have tried to estimate the weight of these factors (Guo and Bhat, 2001).

Understanding the factors which determine household location decisions is essential if the LUTI models are going to realistically simulate the repercussions of introducing different policies and projects into an urban system. These types of long term choices made by households, condition many of the characteristics of urban systems related both to the demand for transport (trip generation/attraction) and for other distance dependent public services (O'Sullivan, 2007).

The studies based on Random Utility Theory (RUT) stand out among research found in the empirical literature which has tried to determine the most important factors influencing residential location. In fact, RUT based research has become the

* Corresponding author. Tel.: +34 942202262; fax: +34 942201703.

E-mail address: dellolio@unican.es (L. dell'Olio).

standard econometric tool both for estimating the determinant factors on residential location as well as in the simulation of these choices within a LUTI system (Pagliara and Wilson, 2010). The application of discrete choice models in the field of residential choice started in the 1970s in the work of Quigley (1976), Lerman (1976) and, from a theoretical point of view, McFadden (1977). These pioneering studies were the forerunners of later studies which have tended to concentrate on specific aspects of residential choice such as its connection with transport choices (Anas, 1981; Eliasson and Mattsson, 2000; Pinjari et al., 2011), the interdependence between the choice of where to reside and the place of work (Anas, 1981; Eliasson and Mattsson, 2000; Pinjari et al., 2011) or the role of accessibility to certain opportunities in the location decisions (Waddell, 1993; Waddell et al., 2007). This article is framed mainly within the latter line of work and, more specifically, the relationship between journey time to work, accessibility to employment and residential location.

Residential choice models can, therefore, help in estimating both the more important factors influencing location decision making, as well as the trade-offs households are faced with. Nevertheless, their use has been criticised (Chen et al., 2008) in that most of the models proposed follow a Multinomial Logit (MNL) specification which is based on the Independence of Irrelevant Alternatives hypothesis (IIA). Although this assumption may be considered correct in other choice contexts it is thought to be inadequate in the field of residential choice. Various authors (Hunt et al., 2004) have argued for the need to estimate using models which consider the existence of spatial correlation between alternatives and, therefore, more complex substitution patterns. A simple substitution pattern as assumed in the MNL model ignores the fact that some places, especially those which are closest, could be better alternatives to the rest in the residential choice context of a certain household. Further criticism concentrated on questioning the validity of the assumption that residential choice is exogenous to the choice of work place (Waddell, 1993). Given that there could be a case where the choice of work place is conditioned by a previous residential choice or that both decisions are made simultaneously, some authors have proposed the use of models which consider both types of choice at the same time.

This article presents the specification and estimation of a series of choice models for residential zone and residential zone together with zone of employment. These models will be estimated in order to check the hypothesis, in accordance with basic economic theory, that the locations with better transport conditions and accessibility to employment will have, *ceteris paribus*, greater utility for the agents and, therefore, will have a greater probability of being chosen as residential areas. The transport conditions in a determined area will be characterised by using home-work journey times considering congestion whereas the accessibility to employment will be based in a gravity type indicator (Handy and Niemeier, 1997). Furthermore, in order to consider the possible existence of spatial correlation, residential choice models will be compared with and without considering the existence of spatial dependence between alternatives. An MNL model will be estimated first, using all the available variables, to estimate the weight of the factors relating to transport to the work place and accessibility to employment. Given that the MNL model is based on the IIA hypothesis, a second step was used to estimate a Nested Logit (NL) model with a structure of correlation between alternatives which was consistent with the maximisation of utility for all possible values of the explicative variables. Finally, a Cross-Nested Logit (CNL) model was estimated considering a free structure of correlation between the alternatives. The specification and estimation of the three models presented here: MNL, NL and CNL leads to posing a question of methodological interest in the field of residential choice: do the models considering spatial dependence between alternatives have a significantly better fit to the data?

The estimated models were applied to a case study which was the urban area of Santander (Spain). The results confirmed that journey times to work place had the expected signs and were a statistically significant factor in explaining household locations. Nevertheless, the accessibility indicator was not significant even though it had the positive sign assumed in the hypothesis. The models considering the possible existence of correlation between alternatives, showed, according to the likelihood-ratio test (LR) a goodness of fit significantly better than that of the MNL, both in the case of the single choice models dealing with residential zone and, at a higher level of confidence, in the choice models dealing with residential zone together with work place location.

The article also provides empirical evidence, in the context of a medium sized city, showing that the cost of the journey from home to work continues to be a statistically important factor when explaining household residential location choices. Previous research also supports this conclusion although the data came from larger metropolitan areas (Dallas–Fort Worth, Puget Sound, etc.) which have many notable differences to medium sized cities (Vernon, 1997). This conclusion is supported with the use of simpler discrete choice models (MNL) as well as by using those that consider spatial correlation between alternatives (NL and CNL).

This work is structured in the following way. Section 2 presents a review of the state of the art in the field of residential location modelling. Section 3 describes the methodology followed and the formulation of the MNL, NL and CNL models. Section 4 presents the application of the methodology to the urban area of Santander and the specification and estimation of the models in order to check the proposed hypotheses. Finally, the conclusions coming from this work are presented in Section 5.

2. Bibliographic review

Work concentrating on the modelling of residential location already has a long history as an independent area of study. There are currently various traditional fields of research that can be identified. Starting from the classification proposed by Pagliara and Wilson (2010), the present article differentiates three broad approaches: based on urban economic theory, based on spatial interaction theory and the approach based on econometric modelling.

The origins of location theory derived from urban economic theory can be traced back to the classic work of von Thünen (1826) on agricultural land use and rents. This pioneering work later became the basis for the creation of the distribution theory of land use and rents in urban areas mainly proposed in the works of Alonso (1964), Muth (1969) and Mills (1972). In the case of residential land use, the nucleus of the theory rests in the modelling of certain trade-offs facing households when choosing where to live. According to the classic theory, the most important trade-off is formed by the dilemma between choosing lower transport costs to the Central Business District (CBD) (where the hypothesis assumes all the jobs are concentrated) and the space occupied by the household. This dilemma can be mathematically modelled by using a bid rent function which decreases as distance to the CBD increases. Following the Alonso–Muth–Mills model, urban location theory has continued to adapt to the ever increasing complexity of urban systems by expanding the classic theory. An excellent systematisation of research can be found in Fujita (1989). It should be pointed out that various LUTI models have also been derived from applied economic–urban theory such as the Penn–Jersey model developed by Herbert and Stevens (1960) from the work of Alonso.

The second of the traditional areas in which the problem of urban residential location has been addressed has been based on spatial interaction theory. This theory appeared at the end of the 19th century to try and explain the regularities found in spatial population flows. The theory establishes an analogy between the movements of people and the attraction of physical objects using universal gravity law. Relevant work in this line can be found in Reilly (1931), Hoyt (1939), Zipf (1949), Isard (1956) and Hansen (1959) whose work is considered to be the beginning of modern spatial interaction theories. However, it was Wilson (1970) who provided a new interpretive framework to the theory of spatial interaction by overcoming the gravitational analogy and substituting it with a probabilistic paradigm based on the maximisation of the entropy of the studied system. Currently, spatial interaction theory under Wilson's paradigm is considered to be a realistic basis for making residential location predictions using optimisation techniques with known destination constraints.

The third line of investigation in the field of residential choice is derived from econometric modelling and, more specifically, discrete choice models. This type of model has proved to be very useful in the field of applied simulation, for making predictions about residential choices made by households about dwelling or residential area. Work developed in the 1970s by Quigley (1976) and Lerman (1976) can be considered as pioneering in this type of modelling, while research carried out by McFadden (1977) signified a considerable advance in this type of study by addressing both the problem of applying the MNL model to choice groups containing many alternatives and to the consideration of the possible correlation between alternatives by introducing the Generalized Extreme Value (GEV) family of models.

As indicated in the introduction, three major lines of investigation can be differentiated in the application of discrete choice models in the field of residential location. A lot of research has taken place relating household location decisions with transport choices. Lerman (1976) specified a choice model which connected residential location, mode of transport to work place and car ownership based on a household survey carried out in the urban area of Washington, DC. The model estimated by Lerman was one of the first to use logit techniques in simulating location choices as a real alternative to traditional techniques based on spatial interaction theory. Anas (1981) applied a Nested Logit choice model connecting location with mode of transport using aggregated data to the metropolitan area of Chicago. The estimations made by Anas showed that the use of aggregated data based on small areas was practically analogous to the use of disaggregated data. In more recent research Eliasson and Mattsson (2000) specified a choice model which connected residential location, journey frequencies, destination and mode of transport. The model developed by Eliasson and Mattsson demonstrated the existence of heterogeneity between destinations as well as the agents within a coherent microeconomic structure similar to that derived from a standard NL model.

Work involving the relationship between accessibility conditions and residential choice has formed another of the main streams of research. Noteworthy research by Guo and Bhat (2001) considered measuring accessibility to green zones, schools, employment opportunities and other public services in a household survey performed in the urban area of Dallas–Fort Worth (USA). The authors showed that accessibility to employment did not appear to be an explanatory factor for residential choice except in the case of highly educated workers. The other measures of accessibility to opportunities which were either leisure or commercial were significant as explanatory variables for making residential choices. Srour et al. (2002) calculated accessibility indexes to be used as explanatory variables in a residential choice model also applied in the Dallas–Fort Worth area. The indicators of accessibility to employment, retail premises and public park space were shown to be significant in the resulting residential choice model and accessibility to employment was shown to be more important than the measures of accessibility to other opportunities. Chen et al. (2008) estimated a MNL model with indicators of accessibility to employment, open spaces, retail areas and leisure opportunities in a panel of households in the metropolitan area of Puget (USA). The authors found that the journey distance to work was a significant factor in residential location. They also underlined the existence of a clear trade-off between journey distance to work and accessibility to open spaces which led them to conclude that households only chose more densely populated areas when they were considerably closer to their places of work.

Although less numerous, another field of research has paid attention to the interrelationship between residential and work place location choices made by households. As shown by the Alonso–Muth–Mills type of models based on the assumption that the urban system being studied has a mono-centric nature with all employment being located around the CBD. This assumption has been one of the more criticised within classic economic theory mainly because of the growing tendency of urban systems in presenting various sub-centres and a growing dispersion of places providing employment (Glaeser, 2008). Within the framework of discrete choice studies, Waddell (1993) specified a choice model addressing place of work, place of residence and type of tenancy based on a household survey in the Dallas–Fort Worth area. Given the large number of

possible alternatives in the household choice group, the specified model was based on a survey of alternatives which followed the methodology established by McFadden (1977). The MNL and NL models estimated by Waddell demonstrated that the grouped choice specification of work place and residential location fit the data better. Waddell et al. (2007) later examined the interdependence between the residential choices and the place of work in the urban area of Puget. To address the problem of the high number of alternatives the authors used a methodology which analysed the sequence of residential and work place location choices by using latent market segmentation. The results showed that this methodology improved the fit of the residential and employment location choice models estimated separately.

One of the most frequently repeated criticisms made against residential choice models based on the use of Multinomial Logit type specifications has been their inability to address spatial correlation between alternatives based on the supposition of IIA. Haynes and Fotheringham (1990) pointed out that aspects such as spatial aggregation or spatial contiguity could produce correlation between alternatives and, therefore, substitution patterns which went against the supposed IIA. The solution to this problem has appeared with the specification of models which allow for more complex substitution patterns. The GEV family of models developed from the work of McFadden (1977), notably the NL model, has provided an initial solution to this problem. However, the NL model still requires the analyst to specify a priori, a correlation structure (Hunt et al., 2004; Pellegrini and Fotheringham, 2002). Bhat and Guo (2004) proposed a model derived from the GEV family which they called Mixed Spatially Correlated Logit (MSCL) for the case of spatially correlated alternatives. This model is based on a special case of the Generalized Nested Logit (GNL) model as originally formulated by Wen and Koppelman (2001) with the incorporation of a distribution of mixes to capture the heterogeneity of tastes between individuals. Bhat and Guo applied the model to a database with the residential choices of a series of households in the metropolitan area of Dallas–Fort Worth. The results of the estimation revealed the importance of the home-work journey costs along and the employment and retail accessibility indicators in choosing where to live. The consideration of the spatial correlation between alternatives improved the fit of the MSCL model over a MNL model. Sener et al. (2011) presented a Generalized Spatially Correlated Logit (GSCL) model to analyse residential choice with correlation between alternatives. This type of model was derived from the GEV family and was based on the MSCL model developed by Bhat and Guo with a more flexible correlation structure between alternatives in which the assignment parameters were considered as a function of a series of zonal characteristics. The GSCL model was applied to a sample of households in the metropolitan area of San Francisco and showed a better fit than the MNL and MSCL models.

3. Application of discrete choice techniques to modelling residential choice

This section summarises the modelling structure used based on the different models within the logit family. The main features of the more basic MNL model will be described first. This will be followed by introducing the NL and CNL models which allow modelling the existence of correlation between alternatives.

3.1. The Multinomial Logit model

Within the family of discrete choice models, the MNL model is the one that has been applied on most occasions in the field of residential choice because of its greater flexibility and simplicity of estimation. The overall utility U of a residential alternative i for a household n can be separated into two components: a systematic utility V_{in} and a random utility ϵ_{in} :

$$U_{in} = V_{in} + \epsilon_{in} \tag{1}$$

where V_{in} takes the form:

$$V_{in} = \sum_k \beta_k X_{ikn} \tag{2}$$

where parameters β_k are to be estimated and where variables X_{ikn} can refer to zonal characteristics in the case of residential choice models, or even in terms of the interaction between the socio-demographic characteristics of the households n and the characteristics of zones i . In household choice models and, generally, in any discrete choice model with alternatives having particular characteristics at an individual level, a series of specific constants can be specified for each alternative. These constants capture the effect of the mean of all the factors which are not observed by the explanatory variables (Ben-Akiva and Bierlaire, 1999). However, in residential choice models, given the aggregated character of the alternatives and the fact that they present characteristics which are identical for all the individuals, specific constants cannot be introduced in the $i - 1$ utility functions. This is because the generic parameters of the different zonal variables may get confused with the specific constants in the estimation. It is still possible though to include dummy variables within the utility functions that represent the influence of large residential districts. However, this practice is not overly recommended because it makes the model less sensitive to simulating the introduction of new policies (Waddell, 2010). Finally, the probability that a household n chooses to live in zone i can be written as:

$$P_{in} = \frac{e^{\lambda V_{in}}}{\sum_{j \in C} e^{\lambda V_{jn}}} \tag{3}$$

where λ is a scaling parameter and C is a choice group.

3.2. Models of the GEV family: NL and CNL

In the context of residential choice models, although the MNL model is easily estimated even in the presence of large choice groups, the IIA property is present. As mentioned earlier, even though this property may be acceptable in multiple choice concepts, it is doubtful that it can be applied to choices that have a strong spatial element which is the case of making a residential choice. So the application of a MNL model to a group of residential choices may lead to the estimation of biased or inconsistent parameters.

The models of the GEV family proposed by McFadden constitute a series of specifications which allow a variety of substitution patterns between alternatives (Train, 2009). All the models in this family share the property that the portions of unobserved utility of all the alternatives distribute together as a Generalized Extreme Value. In the case of the more widely used model from the GEV family, the NL model, the alternatives can be grouped into nests. Within each nest the IIA property is maintained, which is not true for the alternatives of different nests. In the present study the probability of household n choosing a residential zone i is given by:

$$P_{in} = \frac{e^{\lambda_k V_{in}} \left(\sum_{j \in C_k} e^{\lambda_k V_{jn}} \right)^{\lambda_k / \lambda_k}}{\sum_{\ell=1}^K \left(\sum_{j \in C_\ell} e^{\lambda_\ell V_{jn}} \right)^{\lambda_\ell / \lambda_\ell}} \quad (4)$$

where the parameter λ_k is a measure of the degree of independence of the unobserved utility among the alternatives in nest k . The value of λ_k should be between 0 and 1 in order for the model to be consistent with the maximisation of the utility for all the values of the explanatory variables. However, the alternatives belonging to each nest should be specified a priori and require, in the case of the residential location simulation, that the grouping of the residential areas be based on the possible common characteristics not captured by the independent variables.

Ben-Akiva and Bierlaire (1999) proposed a new model in the GEV family which they called Cross-Nested Logit (CNL). The CNL model is an extension of the NL model where an alternative is allowed to belong to more than one nest at the same time. Given this characteristic, the CNL model can be applied in the case of choosing a residential zone without the prior need to impose a spatial structure of correlation between alternatives.

According to Bierlaire (2006) there have been various formulations of the CNL model presented in the literature such as proposed by Small (1987) or Vovsha (1997). However the most general were those presented by Ben-Akiva and Bierlaire (1999) and Wen and Koppelman (2001). The formulation of Ben-Akiva and Bierlaire is based on the following function:

$$G(x_1, \dots, x_j) = \sum_k \left(\sum_{j \in C} \alpha_{jk} x_j^{\lambda_k} \right)^{\lambda_k} \quad (5)$$

where k is the index of the nest, λ_k is the scaling parameter associated to nest k and α_{jk} represents the degree alternative j belongs to nest k . The formulations proposed by Wen and Koppelman and Ben-Akiva and Bierlaire are equivalent although Wen and Koppelman proposed the condition $\lambda = 1$, a common norm in the GEV family of models. Bierlaire (2006) has formally demonstrated that this function fulfils the conditions proposed by McFadden (1977) which allow the CNL model to belong to the GEV family. In addition, the α_{jk} parameters should fulfil the following constraint in order to make the model identifiable:

$$\sum_k \alpha_{jk} = 1 \quad \forall j \quad (6)$$

Finally, in a CNL model, the probability that a residential area i will be chosen by a household n is given by:

$$P_{in} = \frac{\alpha_{ik} e^{\lambda_k V_{in}} \left(\sum_{j \in C} \alpha_{jk} e^{\lambda_k V_{jn}} \right)^{\lambda_k / \lambda_k}}{\sum_{\ell=1}^K \left(\sum_{j \in C} \alpha_{j\ell} e^{\lambda_\ell V_{jn}} \right)^{\lambda_\ell / \lambda_\ell}}$$

4. Application of residential choice models to the metropolitan area of Santander

4.1. Demographic and socioeconomic characteristics of the study area

Santander is a medium sized city located on the north coast of Spain. The city is also the capital of the region of Cantabria. The current population of the city is about 182,700 inhabitants in the urban nucleus; however the overall metropolitan area contains more than 260,000 residents.

The distribution of the population in the study area has dramatically changed recently with a striking increase in the number of people living in the peripheral municipalities during the period 2001–2009 coinciding with a very low population

growth of hardly 1% in Santander itself. This supports the hypothesis that the area is becoming more metropolitan and the population is spreading out over a larger area rather than having high concentrations of people in central urban nuclei. This process is normally described as urban sprawl when it is accompanied by the development of low density residential areas (Palomares, 2007).

More recently, the economy of the metropolitan area has gone through a transition towards the service sector which parallels both the regional and national economies. This can be traced back to the changes occurring in the late 1970s starting a process of continual service sector growth which has led to this sector holding the dominant position within the economic structure of the study area. Grouping all the municipalities in the study area together, in 1986 the service sector represented 60.3% of all employment (Nogues, 1990), whereas in 2008 this figure had grown to 71.8%. Removing the municipality of Santander, the service sector continues to be the predominant provider of employment in the rest of the municipalities, with almost 50% of overall employment. This percentage is slightly above the employment figures provided by secondary industries (manufacturing and construction), which supports the notion of the spread of deindustrialisation throughout the municipalities located around the bay. Although in the 1980s most employment was still provided by secondary industries (47.1% of all jobs) and the primary sector still had an important economic role to play (more than 10% of all jobs), both sectors are currently less significant, especially the primary sector which in 2007 represented less than 1% of overall employment.

The economic structure of the study area can be seen to have undergone important changes over a relatively short recent time scale. These changes can be summarised as:

- Economic activity has definitely moved towards the service sector.
- Although manufacturing still exists (with an important contribution from construction in many municipalities) its presence has been reduced.
- Primary sector industries have lost so much activity that they can be regarded as practically marginal. This is more noteworthy in the eastern municipalities where there has always been a tradition of dairy farming.
- From a spatial point of view, the traditional zonal division into the capital as the central provider of services, the north western zone containing the industrial municipalities and the eastern zone occupied by dairy farming and leisure activities has to a certain extent been redrawn. There is currently a dominance of tertiary sector industries in almost all the municipalities.
- Activity continues to be highly concentrated in the municipality of Santander although a certain trend towards some diffusion of these services can be seen in favour of other metropolitan municipalities.
- The creation of large out-of-town shopping malls has strengthened this diffusion in terms of retail trade by drawing a large number of shoppers outside the commercial centre of Santander towards the suburbs whilst at the same time attracting people from other neighbouring areas.

4.2. Available data

A sample of 534 households was used. This sample came from a simple random sample mobility survey carried out in 2008 on a section of the population from households in the urban area. Given that the study concentrates on the influence of factors relating to accessibility and transport to work, the chosen sample is made up of households that have only one member with a salary. This choice removes the problem associated with the number of workers present in each household when specifying the model (Waddell, 1996).

In fact, of the 534 surveyed households, 396 had at least one worker and of those 275 had only one worker. The geographical distribution of the properties according to the number of workers they hold does not follow any particular pattern and can be considered to be quite disperse, it has been assumed that the hypothesis of including one-worker households does not affect the end result of the study.

The survey collected information on household location, their basic characteristics (number of members, income level, etc.) and their revealed mobility preferences detailed in a journey diary. The postal address was used to code each observation into a geographical information system (GIS). The introduction of the data into the GIS meant that certain variables could be obtained about the surroundings and locations of each one of the households by cross-referencing the point location data of each home with the variables present in the land use zoning used. This zoning divided the municipal area into a total of 26 zones. Given the limited size of the overall study area it was thought that 26 zones with an median area of 0.33 km² fulfilled the condition imposed by Anas (1981) who worked with zoning in squares of 0.65 km².

The variables contained in the database are presented in Table 1 and can be classified into three types: variables relating to the environmental conditions in the zone, variables relating to the accessibility and transport to work conditions, and interaction variables. The summary of the descriptive statistic of the variables is given in Table 2.

The choice of explanatory variables was strongly conditioned by the available sources. An attempt was made to make the independent variables take into account the following the three main aspects which are generally considered to be a key part of residential choice:

- The environmental conditions in the residential area: this group contains “employment”, “foreigners”, “prestige”, “house price”, “learning” and “dwellings”.

Table 1
Description of the explanatory variables used.

Variable	Description	Type
Accessibility	Accessibility to employment zone	Transport
Journey time	Journey time between residential zone–employment zone	Transport
Residence/work	Dummy = 1 where the zone of residence and zone of work coincide	Transport
Waiting time	Average waiting time at public transport stops in the zone	Transport
Employment	No. of jobs in the zone	Environmental
Foreigners	No. of none EU foreign residents present in the zone	Environmental
Dwellings	Natural log of the No. of dwellings in the zone	Environmental
Prestige	Dummy for a zone with special prestige	Environmental
House price	Average house price for the zone	Environmental
Learning	Learning centres at least 1000 m from zone centroid	Environmental
High Income	Dummy for income >2500	Interaction
Age	Age	Interaction
Work	No. Worker per household	Interaction

Table 2
Descriptive statistic.

Variable	Mean	Standard Deviation	Minimum	Maximum	Measurement unit
Accessibility	29.34	7.42	10.02	48.05	–
Journey time	13.51	1.88	10.10	18.83	Minutes
Residence/work	0.07	0.25	0.00	1.00	–
Waiting time	10.51	0.78	8.50	12.11	Minutes
Employment	2602.26	2737.53	520	11,359	No. Jobs
Foreigners	461.12	224.36	108.00	843.00	No. Foreigners
Dwellings	2650.95	567.20	1072.00	3804.00	No. dwellings
Prestige	0.09	0.29	0.00	1.00	–
House price	287609.60	126695.66	174920.83	930899.18	€
Learning	2.22	1.70	0.00	7.00	No. Schools
Income	2.90	0.79	1.00	4.00	–
Age	43.85	11.91	16.00	82.00	Years
Work	1.8	0.68	1.00	4.00	No. Workers per household

- The accessibility and transport conditions of each residential area: these are the most interesting for the objectives of this research. This group may include: “accessibility”, “journey time”, “Residence/work” and “Waiting time”.
- The demographic characteristics of the population. Household income has been taken as particularly relevant because budgetary constraints are usually an important conditioning factor on location choice.

Furthermore, other variables referring to the socio-demographic characteristics of households were considered, even though they were not thought to be theoretically relevant or did not turn out to be significant in the process of specifying the residential choice models: these were the age of household members (“Age”), and the number of workers in the household (“Work”).

Interactions consider how the environmental and transport conditions differentially affect households according to their income levels. A dummy variable was used to consider these differential effects in households with income levels above the third quartile (above 2500 €) on the variables “accessibility”, “journey time”, “foreigners”, “dwellings”, “prestige”, “house price”, “learning” and “waiting time”.

The variables “employment”, “foreigners”, “dwellings” and “learning” were established from different statistical operations made by the Spanish Office of National Statistics (INE) and by the Cantabrian Institute of Statistics (ICANE). The “learning” variable can be considered as an accumulated opportunities type of accessibility indicator to centres of primary and secondary education in each of the zones. The average price per zone (“house price”) is collected from a series of real estate sources which represent asking prices on a sample of 845 properties in the study area. The prestige of the zones (“prestige”) depended on their belonging to the city centre or the better known neighbourhoods on a regional or even national scale (e.g. the El Sardinero neighbourhood).

The variables considered to be more relevant to the objectives of this study are those which refer to the accessibility and transport to work conditions: “accessibility”, “journey time”, “Residence/work” and to a lesser extent, “waiting time”. The average waiting times in the zones were calculated from the data provided by the municipal public transport service. The initial hypothesis was that this variable would have a negative effect on a zone’s utility, even though it did not account for all the journey to work costs using public transport it was not thought to be fundamental for the goals of this study. The “journey time” variable represents access time in minutes from the head of the household’s residential zone to their employment zone. This variable has been calculated with a transport model using the real morning rush hour traffic flows. The times that were included in the database therefore take congestion into account and this variable is expected to present

a parameter with a negative sign in the models. “Residence/work” represents a dummy variable taking a value of 1 if the residential zone and employment zone coincide. This variable is expected to have a positive sign which signifies that the households tend to assign greater utility to those zones where their work place is located *ceteris paribus*. Finally “accessibility” is a Hansen type gravity indicator of accessibility to employment which considers the possible multicenter nature of the urban area. The indicator was calculated from the employment data present in the zones and the following expression was used (Coppola and Nuzzolo, 2011):

$$Acc(o) = \sum_i [\exp(\alpha_2 \cdot Cost(o, d_i)) \cdot jobs(d_i)^{\alpha_1}] \tag{7}$$

where *Cost* is a measure of the journey cost between origin *o* and destination *d_i* by car calculated using a transport model which considered congestion at morning rush hour. *Jobs (d_i)* are the number of employments present in the destination zone *d_i* and α_1 and α_2 are parameters to be estimated. The parameters may be calculated by linearisation (8) taking logarithms to both sides of the expression. The parameters are then estimated by ordinary least squares taking the transport flows generated by each zone as the dependent variable for the accessibility conditions. The parameter α_1 presented an estimated value of 0.26 while α_2 had a value of -0.12. The estimation had a *R*² goodness of fit of 0.7. The parameter of this variable was expected to have a positive value meaning that households tended to give greater utility to the zones with more employment close by.

4.3. Multinomial Logit model for choosing residential zone and residential zone considering employment zone

This subsection presents the estimation of two residential choice models using MNL. The first of the estimated models (MNL-1) considers the location of the place of employment as an exogenous factor. Under this supposition the group of choices available to households is limited to a total of 26 alternatives. The second model to be estimated is a joint logit model which considers the joint choice of place of work and place of residence (see MNL-2 in Table 3 with the *p*-values of the significance of the parameters in brackets). In this case the expression of utility is formulated as:

$$U_{rwn} = V_m + V_{wn} + V_{rwn} + \epsilon_{rwn} \tag{8}$$

where *U_{rwn}* is the overall utility of choosing zone of residence *r* and place of work *w* for household *n*. *V_m* is the systematic utility of choosing zone of residence *r* for household *n*, *V_{wn}* is the systematic utility of choosing zone of work place *w*, *V_{rwn}* is the systematic utility of the specific combination of zone of residence and work *rw* and ϵ_{rwn} is the random component of the alternative’s utility.

As stated before, some authors have pointed out the need to consider that the choice of residential zone is not independent of the choice of work place zone, because if it is not considered it may reduce the goodness of fit of the model and bias

Table 3
Parameters estimated for the MNL residential location models.

Type	Variables	Interaction	MNL-1. Choice of residential zone	MNL-2. Joint choice of residential and work location zone	
Transport	“Accessibility”	–	.005 (.56)	.006 (.53)	
	“Accessibility”	“High Income”	–.016 (.38)	–.016 (.38)	
	“Journey time”	–	–.102 (.01)	–.103 (.01)	
	“Journey time”	“High Income”	–.000 (.92)	–.007 (.92)	
	“Residence/work”	–	.235 (.29)	.233 (.29)	
	“Waiting time”	–	–.132 (.10)	–.137 (.09)	
	“Waiting time”	“High Income”	–.013 (.94)	–.021 (.91)	
Environmental	“Employment”	–	–	.000 (.00)	
	“Foreigners”	–	–.000 (.04)	–.000 (.04)	
	“Foreigners”	“High Income”	–.001 (.12)	–.001 (.13)	
	“Dwellings”	–	1.39 (.00)	1.39 (.00)	
	“Dwellings”	“High Income”	1.49 (.07)	1.42 (.08)	
	“Prestige”	–	–.897 (.00)	–.934 (.00)	
	“Prestige”	“High Income”	1.90 (.00)	1.93 (.00)	
	“House price”	–	–.000 (.04)	–.000 (.05)	
	“House price”	“High Income”	–.000 (.57)	–.000 (.55)	
	“Learning”	–	–.087 (.05)	–.090 (.04)	
	“Learning”	“High Income”	.242 (.00)	.246 (.00)	
		Null Log-Likelihood		–1739.82	–3479.64
		Log-Likelihood		–1658.58	–3274.40
	LR test Null		162.48	410.48	
	No. Alternatives		26	676	
	N		534	534	

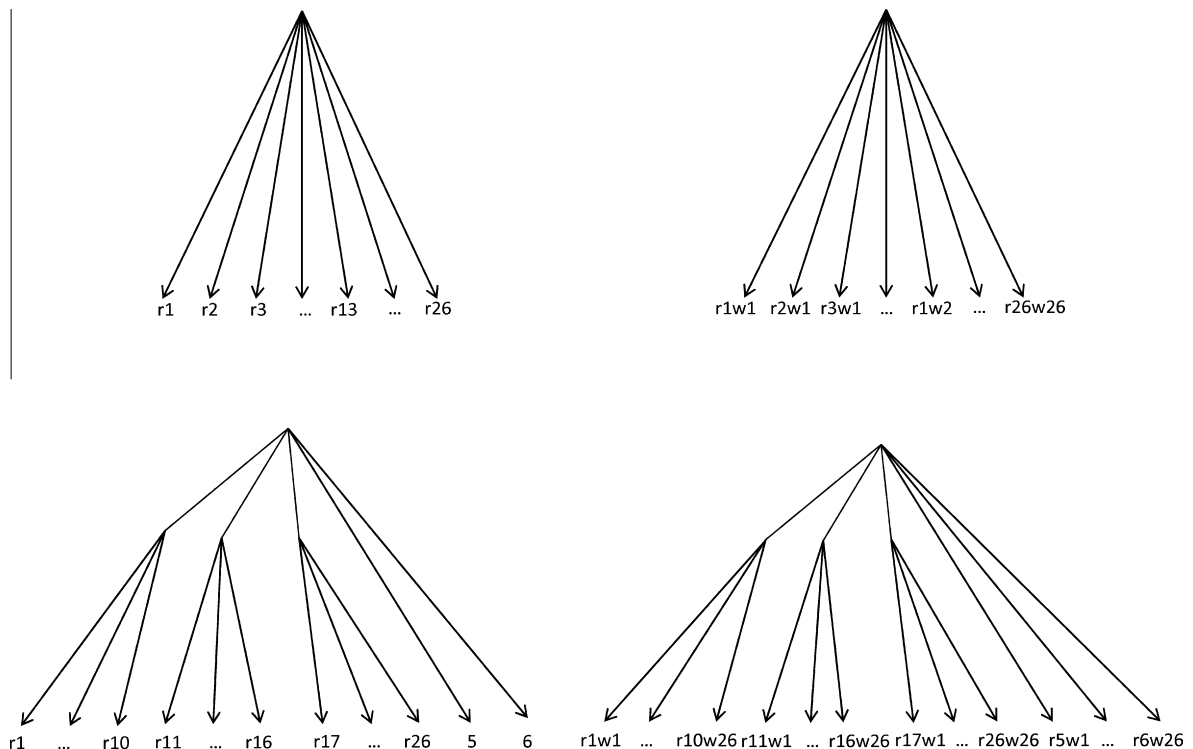


Fig. 1. Nested structures of the MNL-1 models (top left), MNL-2 (top right), NL-1 (bottom left) and NL-2 (bottom right).

the estimated parameters (Waddell, 1993). Therefore, in this model the households present a total of $26 * 26 = 676$ alternatives, which is an ample group of choices (see Fig. 1 to compare the structure of the models). Previous research with large choice groups has used the random sampling of alternatives technique proposed by McFadden, even though this can only be used in the case of MNL models unless additional corrections are used (Lee and Waddell, 2010). However, the current software available for estimations (e.g. Biogeme (Bierlaire, 2003)) enable working with large choice groups but with rather high estimating times.

The residential choice model MNL-1 estimated with a choice group of 26 alternatives presented a log-likelihood in convergence of -1658.58 . From the transport related variables, the “accessibility” parameter was not significant neither for the overall group of households nor for the households with incomes of over 2500 €, although it did present a theoretically believable sign. Nevertheless, the coefficient of “journey time” did turn out to be significant in the sample group of households and it had a negative sign which agreed with the initial expectations. The parameter of the variable referring to “waiting times”, was significant at a 90% confidence level and had a negative sign meaning that households considered that increased waiting times for public transport reduced utility. Finally although the “Residence/work” variable produced a positive parameter it was only significant at a 75% confidence level.

The coefficients relating to the environmental variables such as “foreigners”, “dwellings”, and “house price” had the expected signs and, in general, were significant. In the cases where they presented theoretically incoherent signs in the population group such as in the cases of “prestige” or “learning”, they were as expected for the higher income group of households. So, given that the data are based on a revealed preferences survey, only the higher income households tend to locate in areas with a high number of schools or in prestigious zones because their parameters, once added to the parameters for the overall population, continue to be positive.

The MNL-2 joint choice model for residential zone and work place zone generally presented similar parameters for all the variables to those estimated using MNL-1. An additional variable (“employment”) was introduced for modelling the choice of work place. “Employment” variable was clearly significant adding a utility of 0.0001 to the alternatives for each additional job.

Both models presented a significant fit with respect to the equiprobable model according to the LR test even though the result was higher when places of residence and work were modelled together (410.48 versus 162.48).

4.4. Models considering spatial correlation between alternatives: NL and CNL

This section introduces the estimations made for the models considering the existence of spatial correlation between alternatives. Two NL models were initially estimated, the first of these (NL-1) considered, as in the case of the MNL models,

the choice of employment zone as being exogenous to the choice of residential zone, whereas the NL-2 model again considers the choices of residential zone and employment zone as a joint process.

Various specifications were estimated to consider different structures of prior correlation between zones by grouping them into different nests. The specification that was finally selected for NL-1 (see Fig. 1 and Table 4) had three nests plus two alternatives hanging directly from the root nest. The overall model should fulfil certain conditions in order to be consistent with Random Utility Theory (Train, 2009). Firstly, one of the scaling factors has to be established for it to be estimated. In this case an upper level normalisation of $\lambda = 1$ was chosen. Secondly, the estimated nest parameters should fulfil the condition that $\lambda_k > 1$ so the coefficient λ/λ_k is between 0 and 1 thereby assuring that an increase in the utility of any of the alternatives from the nest increases the probability of choosing it as a whole.

Finally, one of the proposed nesting structures was chosen with the alternatives grouped into three main nests. These nests divided the study area into three large zones (see Fig. 2). Firstly, nest A grouped alternatives 1–4 and 7–10 corresponding to the central and north-eastern areas of the city, a zone that generally had a higher residential status. Nest B grouped together the alternatives 11–16 belonging to the neighbourhoods around the city centre which were considered as residential areas with a generally lower status. Finally, nest C grouped together alternatives 17–26 corresponding to the western part of the studied area, with a generally varied nature consisting of urban and peri-urban residential areas, as well as industrial zones and transport infrastructure installations. Two alternatives, 5 and 6 were grouped into nest D with a fixed parameter of 1 and therefore directly connected with the root nest. This was because different specifications showed that they had little correlation with the other areas or between themselves. In the NL-2 model, the same correlation structure was replicated between the alternatives grouping the joint choices of residential zone and work place zone rw according to the residential zone.

The parameters estimated using NL-1 and NL-2 were very similar to those obtained from the MNL specifications. Among the environmental variables only the “learning” coefficient became non-significant in the NL models, even though for the higher income households (“learning” interacting with “High income”) the greater number of centres of learning was a positive and significant factor in choosing location. The “waiting time” parameter was the least significant from the variables relating to accessibility and transport to work place, although it kept its negative sign in the overall household survey. Furthermore, the “waiting time” interacting with “high income” parameter presented an unexpected positive sign, even though it was clearly not significantly different from zero. The number of jobs present in the zone variable, “employment”, once again returned a clearly significant parameter. The remaining interesting variables “accessibility”, “journey time” and IN did not show any changes in their parameters nor in their signs and practically not even in magnitude and significance. The coefficients of the A, B and C nests were greater than 1 although they were not significant in all cases. In particular, nest

Table 4
Estimated parameters for the NL and CNL residential location models considering the existence of correlation between alternatives.

Type	Variables	Interaction	NL-1. Choice of residential zone	NL-2. Joint choice of residential and work place zone	CNL-1. Choice of residential zone	CNL-2. Joint choice of residential and work place zone	
Transport	“Accessibility”	–	.005 (.51)	.007 (.42)	.006 (.33)	.002 (.80)	
	“Accessibility”	“High Income”	–.013 (.39)	–.015 (.38)	–.003 (.79)	–.008 (.62)	
	“Journey Time”	–	–.089 (.01)	–.093 (.02)	–.068 (.02)	–.064 (.08)	
	“Journey time”	“High Income”	–.000 (.99)	–.009 (.89)	–.000 (.99)	–.014 (.82)	
	“Residence/work”	–	.214 (.26)	.236 (.26)	.203 (.14)	.317 (.13)	
	“Waiting time”	–	–.097 (.18)	–.083 (.30)	–.069 (.32)	–.127 (.14)	
	“Waiting time”	“High Income”	.048 (.76)	.027 (.88)	.061 (.65)	–.047 (.79)	
	Environmental	“Employment”	–	–	.000 (.00)	–	.000 (.00)
“Foreigners”		–	–.000 (.04)	–.000 (.04)	–.000 (.13)	–.000 (.15)	
“Foreigners”		“High Income”	–.001 (.12)	–.001 (.13)	–.000 (.07)	–.001 (.06)	
“Dwellings”		–	1.28 (.00)	1.46 (.00)	.531 (.07)	1.26 (.00)	
“Dwellings”		“High Income”	1.15 (.08)	1.34 (.09)	.685 (.13)	1.52 (.06)	
“Prestige”		–	–.681 (.01)	–.685 (.02)	–.339 (.12)	–.921 (.00)	
“Prestige”		“High Income”	1.76 (.00)	1.90 (.00)	1.19 (.00)	1.59 (.00)	
“House Price”		–	–.000 (.01)	–.000 (.01)	–.000 (.00)	–.000 (.16)	
“House price”		“High Income”	–.000 (.50)	–.000 (.40)	–.000 (.53)	–.000 (.83)	
“Learning”		–	–.053 (.17)	–.053 (.23)	.000 (.98)	–.109 (.02)	
“Learning”		“High Income”	.215 (.00)	.236 (.00)	.152 (.00)	.191 (.01)	
NESTA				1.24 (.11)	1.07 (.09)	3.97 (.01)	1.82 (.00)
NESTB				1.28 (.05)	1.10 (.02)	1.85 (.05)	1.21 (.03)
NESTC				1.11 (.28)	1.03 (.35)	1.27 (.00)	1.00 –
NESTD				1.00	1.00	1.00	1.00
Null Log-Likelihood				–1739.82	–3479.64	–1739.82	–3479.64
Log-Likelihood				–1654.73	–3270.17	–1626.64	–3070.01
LR test Null			170.17	418.94	226.35	819.26	
LR test MNL/NL			7.69	8.46	56.18	400.32	
No. Alternatives			26	676	26	676	
N			534	534	534	534	

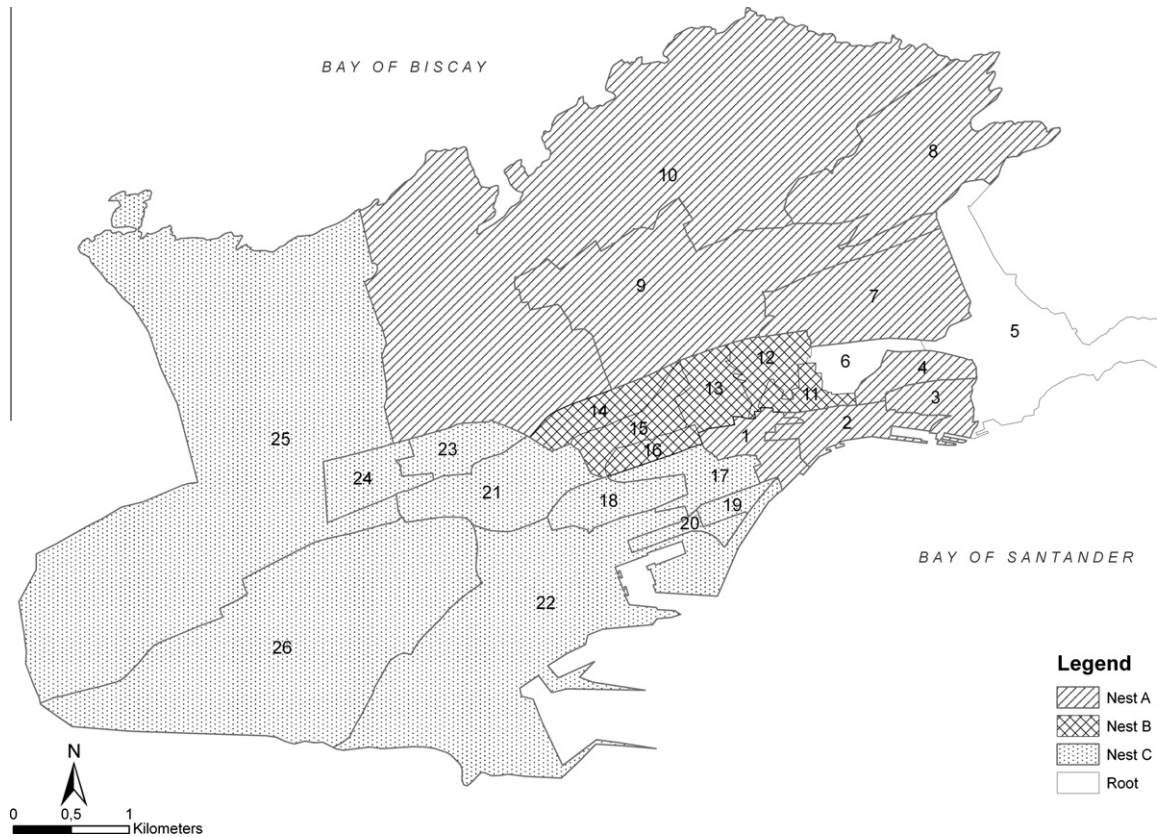


Fig. 2. Land use zoning used in the study area.

C was only significant at a 72% confidence level and at 65% for the residential location and joint residential and work place location models, respectively.

Both the residential location model and the joint location model had clearly better and more significant fits than the respective equiprobable models. Furthermore, if their fit is compared to the one obtained for the MNL models using the LR test, the residential location model offers a significantly better fit at a 90% confidence level, while the joint location model presented a better value in the test (8.46) at a confidence level greater than 95%.

Finally, CNL models were estimated with identical specification to the NL models but without the need for prior detailing a spatial structure of correlation between alternatives. Nevertheless, in order to limit the number of parameters to be estimated in the models, a technique that had been previously used in other studies (Bhat and Guo, 2004) was applied to establish the condition that a zone could only belong to those nests which presented common borders. Only alternatives 5 and 6 continued to hold direct connected with the root nest.

The models had long calculation times because of the many parameters to be estimated, especially in the case of the CNL joint choice model of residential and work place zone. In the case of the joint residential-work place model the software used required several days to calculate the variances–covariances matrix for obtaining the statistical significance (see Table 4).

The CNL-1 residential choice model, showed quite similar parameters to those obtained from the MNL and NL models. Among the environmental variables, the “dwellings” parameter had the expected sign but with a lower magnitude and was only significant at a 93% confidence level. The coefficient of “learning” was clearly not significant even though, the number of schools present in the zones of higher income households (“learning” interacting with “High Income”), did turn out to be significantly different from zero with a positive sign. Among the variables related to accessibility and transport to work not even the parameters of “accessibility” or “waiting time” were significant even though they had a p -value of around 0.3. The parameter of “journey time” however, continued to present the expected negative sign and was clearly significant. Finally, the parameter of the “Residence/work” variable and its significance were similar to those present in the MNL and NL models. In the case of the CNL-2 model the estimated parameters were once again quite similar to those already mentioned above. The “accessibility” variable presented a parameter with a positive sign and a magnitude a little greater than estimated in the MNL and NL models as in the case of the CNL-1 model. The parameters of “journey time”, IN and “waiting time” continued to show signs and magnitudes which were very similar to those estimated using the other models.

In the case of the CNL-1 all the λ_k scaling parameters were significantly different from one, indicating that the CNL specification helped increase correlation inside the nests. In the joint choice model CNL-2, nests A and B presented parameters

significant, while the nest C took a value equal to 1 which is an indication of the non-existence of correlation between alternatives, a fact similar to the present in the NL models.

The goodness of fit of the CNL models was greater than found using the MNL and NL models. The LR test used in CNL-1 against the fit of NL-1 and in CNL-2 against the fit of NL-2, was significantly better at 90% and 95% confidence levels, respectively, in spite of the great number of additional parameters that had to be estimated for the CNL type of models.

5. Conclusions

This article has presented the specification of three types of discrete choice models MNL, NL and CNL considering the choice of residential zone and the joint choice of residential zone and work place zone. The models were specified to estimate the influence of accessibility and transport to work conditions on the choice of residential zone made by a sample of households. The estimation of these models using data from the urban area of Santander also allowed them to be compared to determine if the models that considered the existence of spatial correlation between alternatives presented a significantly better fit to the data.

In considering the effects of the factors relating to accessibility and transport to work, the Hansen type accessibility to work indicator introduced in the specification of the models was generally found not to be significant, even though it always presented a positive sign for the group of households in the survey. On the other hand, the parameter of the variable referring to home-work journey times was clearly significant in all the models at least a level of confidence of 90% with no significant differences between the overall group of households and the group with higher incomes. This result can therefore be considered to be compatible with the results obtained in previous research such as Guo and Bhat (2001) and Bhat and Guo (2004). The “Residence/work” variable referring to the location of work place and residence in the same zone also had a positive sign using all the models even if it was only significant at a confidence level of between 71% and 87%. Finally, although waiting times for public transport had the expected negative sign in most of the models, they were only significant at a 90% confidence level in the MNL models which offered the worst fit. These facts show that, in the sample analysed, journey times to work were definitely an important factor for households in choosing where to live. Furthermore, the lack of significance of the accessibility indicator could be due to part of its effects being captured by the “journey time” and “Residence/work” variables. It can therefore be concluded that, in agreement with the hypothesis proposed at the beginning of this research, lower journey times to work continue to be an important factor when households are deciding where to live.

The environmental variables, considered to be secondary for the purposes of this study, generally presented the expected negative signs in the cases of “foreigners” (an indicator of the existence of a certain degree of spatial segregation) and “house price” and positive in the case of “dwellings”. The parameters of the variables “prestige” and “learning” also behaved differentially according to household income levels: while the prestige of a zone was a strong positive factor for the utility of the alternatives in the case of higher income households, in the case of the overall group of households this variable presented a negative parameter using most of the models. Similar results were found with the parameters of the variable referring to the number of schools.

The joint choice models for residential and work place location zones showed similar parameters to those obtained in the purely residential choice models, including the additional (“employment”) variable, always significant and with the correct sign. Furthermore, the LR test showed that the NL and CNL joint choice models offered a better fit than the MNL at a 95% confidence level. This level of confidence was certainly better than the 90% obtained by the purely residential choice models.

The NL models considering the existence of spatial correlation between alternatives had a better fit with the data than the MNL models, even though in some cases the parameters of the nests were not significant to a high level of confidence. In comparison, the CNL-1 model had a good fit with all the parameters of the nests, significantly different from one and at a confidence level of at least 95%. This fact shows the presence of a certain degree of spatial correlation between the alternatives which was captured more precisely by the CNL model than by the previously specified and closed structure of the NL models. The parameters of the CNL-2 models were similar to those of other models. Moreover the correlation structure allowed getting a better fit to the data even if this was at the expense of rather high estimating times caused by the greater number of parameters. Therefore this type of model looks promising for the future in the field of residential choice as the capacity of software increases to enable estimations to be made with large choice sets and specifications with the presence of a large number of parameters.

Acknowledgements

This study was supported by research funding from the Spanish Ministry of Science and Innovation through the Project E 21/08 – INTERLAND.

References

- Alonso, W., 1964. *Location and Land Use: Toward a General Theory of Land Rent*. Harvard University Press, Cambridge.
- Anas, A., 1981. The estimation of multinomial logit models of joint location and travel mode choice from aggregated data. *Journal of Regional Science* 21 (2), 223–242.
- Barra, T.d.l., 1989. *Integrated Land Use and Transport Modelling: Decision Chains and Hierarchies*. Cambridge University Press, Cambridge, New York.
- Ben-Akiva, M., Bierlaire, M., 1999. *Discrete Choice Methods and their Applications to Short Term Travel Decisions*. Handbook of Transportation Science.

- Bhat, C.R., Guo, J., 2004. A mixed spatially correlated logit model: formulation and application to residential choice modeling. *Transportation Research Part B: Methodological* 38 (2), 147–168.
- Bierlaire, M., 2003. Biogeme: a free package for the estimation of discrete choice models. In: *Proceedings of the 3rd Swiss Transportation Research Conference*, Ascona, Switzerland.
- Bierlaire, M., 2006. A theoretical analysis of the cross-nested logit model. *Annals of Operations Research* 144 (1), 287–300.
- Chen, J., Chen, C., Timmermans, H., 2008. Accessibility trade-offs in household residential location decisions. *Transportation Research Record: Journal of the Transportation Research Board* 2077 (-1), 71–79.
- Coppola, P., Nuzzolo, A., 2011. Changing accessibility, dwelling price and the spatial distribution of socio-economic activities. *Research in Transportation Economics* 31, 63–71.
- Eliasson, J., Mattsson, L.-G., 2000. A model for integrated analysis of household location and travel choices. *Transportation Research Part A: Policy and Practice* 34 (5), 375–394.
- Fujita, M., 1989. *Urban Economic Theory: Land Use and City Size*. Cambridge University Press, Cambridge, Cambridge, New York.
- Glaeser, E.L., 2008. *Cities, Agglomeration, and Spatial Equilibrium*. Oxford University Press, Oxford.
- Guo, J., Bhat, C., 2001. *Residential Location Choice Modeling: Accommodating Sociodemographic, School Quality and Accessibility Effects*. University of Texas, Austin.
- Handy, S., Niemeier, D., 1997. Measuring accessibility: an exploration of issues and alternatives. *Environment and Planning* 29, 1175–1194.
- Hansen, W.G., 1959. How accessibility shapes land use. *Journal of the American Institute of Planners* 25 (2), 73–76.
- Haynes, K.E., Fotheringham, A.S., 1990. The impact of space on the application of discrete choice models. *Review of Regional Studies* 20, 39–49.
- Herbert, J.D., Stevens, B.H., 1960. A model for the distribution of residential activity in urban areas. *Journal of Regional Science* 2 (2), 21–36.
- Hoyt, H., 1939. *The Structure and Growth of Residential Neighbourhoods in American Cities*. Federal Housing Administration, Washington, DC.
- Hunt, L.M., Boots, B., Kanaroglou, P.S., 2004. Spatial choice modelling: new opportunities to incorporate space into substitution patterns. *Progress in Human Geography* 28 (6), 746–766.
- Isard, W., 1956. *Location and Space-Economy*. MIT Press, Cambridge, MA.
- Lee, B., Waddell, P., 2010. Residential mobility and location choice: a nested logit model with sampling of alternatives. *Transportation* 37 (4), 587–601.
- Lerman, S.R., 1976. Location, housing, automobile ownership and mode to work: a joint choice model. *Transportation Research Record* (610), 6–11.
- Lowry, I.S., 1964. *A Model of Metropolis*. Rand Corporation, Santa Monica, Calif.
- McFadden, D.L., 1977. *Modelling the Choice of Residential Location*. Cowles Foundation for Research in Economics, Yale University.
- Mills, E.S., 1972. *Studies in the Structure of the Urban Economy*. Published for Resources for the Future by Johns Hopkins Press, Baltimore.
- Muth, R.F., 1969. *Cities and Housing: The Spatial Pattern of Urban Residential Land Use*. University of Chicago Press, Chicago.
- Nogues, S., 1990. Caracterización socioeconómica y territorial del área de la bahía de Santander. In: *La bahía de Santander: actas de las Jornadas de estudios territoriales de Cantabria*, 116, Santander. Asociación Cantabra de Ciencia Regional.
- O'Sullivan, A., 2007. *Urban Economics*, sixth ed. McGraw-Hill/Irwin, Boston, London.
- Palomares, J.C.G., 2007. *Movilidad laboral en la Comunidad de Madrid*. Ph.D. Thesis. Madrid. Universidad Complutense.
- Pagliara, F., Wilson, A., 2010. The state-of-the-art in building residential location models. In: Pagliara, F., Preston, J., Simmonds, D. (Eds.), *Residential Location Choice*. Springer, Berlin Heidelberg, pp. 1–20.
- Pellegrini, P.A., Fotheringham, A.S., 2002. Modelling spatial choice: a review and synthesis in a migration context. *Progress in Human Geography* 26 (4), 487–510.
- Pinjari, A., Pendyala, R., Bhat, C., Waddell, P., 2011. Modeling the choice continuum: an integrated model of residential location, auto ownership, bicycle ownership, and commute mode choice decisions. *Transportation*, 1–26.
- Quigley, J., 1976. *Housing Demand in the Short Run: An Analysis of Polytomous Choice*. National Bureau of Economic Research, Inc., pp. 76–102.
- Reilly, W.J., 1931. *The Law of Retail Gravitation*. W.J. Reilly, New York.
- Sener, I.N., Pendyala, R.M., Bhat, C.R., 2011. Accommodating spatial correlation across choice alternatives in discrete choice models: an application to modeling residential location choice behavior. *Journal of Transport Geography* 19 (2), 294–303.
- Small, K.A., 1987. A discrete choice model for ordered alternatives. *Econometrica* 55 (2), 409–424.
- Srouf, I., Kockelman, K., Dunn, T., 2002. Accessibility indices: connection to residential land prices and location choices. *Transportation Research Record: Journal of the Transportation Research Board* 1805 (-1), 25–34.
- Train, K., 2009. *Discrete Choice Methods with Simulation*, second ed. Cambridge University Press, Cambridge, New York.
- Vernon, H., 1997. Medium size cities. *Regional Science and Urban Economics* 27, 583–612.
- von Thünen, J.H., 1826. *Der isolierte staat in beziehung auf landwirtschaft und nationaloekonomie*, Jena (C.M. Wartenburg, Trans. (1966)). The Isolated State. Oxford University Press, Oxford, UK.
- Vovsha, P., 1997. Cross-nested logit model: an application to mode choice in the Tel-Aviv metropolitan area.
- Waddell, P., 1993. Exogenous workplace choice in residential location models: is the assumption valid? *Geographical Analysis* 25 (1), 65–82.
- Waddell, P., 1996. *Accessibility and Residential Location: The Interaction of Workplace, Residential Mobility, Tenure, and Location Choices*. Lincoln Land Institute TRED Conference.
- Waddell, P., 2002. UrbanSim: Modeling urban development for land use, transportation, and environmental planning. *Journal of the American Planning Association* 68 (3), 297–343.
- Waddell, P., 2010. Modeling residential location in UrbanSim. In: Pagliara, F., Preston, J., Simmonds, D. (Eds.), *Residential Location Choice*. Springer, Berlin Heidelberg, pp. 165–180.
- Waddell, P., Bhat, C., Eluru, N., Wang, L., Pendyala, R.M., 2007. Modeling interdependence in household residence and workplace choices. *Transportation Research Record*, 84–92.
- Wen, C.-H., Koppelman, F.S., 2001. The generalized nested logit model. *Transportation Research Part B: Methodological* 35 (7), 627–641.
- Wilson, A.G., 1970. *Entropy in urban and regional modelling*. Pion, London.
- Zipf, G.K., 1949. *Human Behavior and the Principle of Least Effort*. Addison-Wesley, Cambridge, MA.