Exploring the Factors that Impact on Transit Use through an Ordered Probit Model: the Case of Metro of Madrid

Laura Eboli, Carmen Forciniti, Gabriella Mazzulla
University of Calabria, Italy
Francisco Calvo
University of Granada, Spain

SUMMARY

The configuration of urban areas is the result of a cyclic relationship between land use and transportation system: the changes in transportation system arrangements influence the localisation of residence and economic activities, as well as the changes in land use affect transportation system characteristics. In this context, by operating on land use, travel demand can be shift from the individual transportation modes to transit systems. In the literature, many conceptual models were proposed to describe the complex relationship between land use and travel behaviour. In addition to spatial variation, the study of travel demand shows the categorical variation of variables.

This work aims to analyse the influence of the categorical variation of variables impacting on transit use. An ordered probit model is proposed for evaluating how transit use depends on variables related to socio-economic characteristics of population, territorial features, accessibility, and transportation system. The study case is Madrid metro network (Spain).

The results show a strong influence of characteristics of population and land use variables on daily trips made using metro system and highlighted the aspects that mainly impact on the choice to travel by metro, providing useful suggestions for shifting people from individual transportation mode to transit systems.

1. INTRODUCTION

Urban areas configuration is the result of the interactions between transportation system and land use. The changes that occur in the transportation system can influence users' behaviour and future localisation of residence and activities. Similarly, the changes in land use can produce modification to the transportation system and users' travel behaviour.

In the last few years, the decline in railway use has favoured the expansion of road mobility and its infrastructures (Nocera et al., 2012). From the user's perspective, private car is perceived as the most comfortable, flexible, and fast compared with the other modes. In addition, the urban space characteristics, as urban sprawl and low density areas, contribute to increase the attractiveness of car as the most convenient transportation mode over all the points of view. The fast increase of the number of trips made by car causes the complication of the air pollution problem and the energy consumption. Simultaneously, it participates to arise of social problems as traffic congestion and decline in quality of life. Consequently, transportation and territorial planning aim to shift travel demand from private car to public transit systems. In order to attract more users, the transport companies can operate

interventions aimed to increase the service quality and to reduce the travel time. On the other hand, the research in the field of land use and transport interaction can focus on the identification of interventions on land use which could positively influence users' travel behaviour. In this context, by operating on land use, travel demand can be shift from the individual transport to transit systems.

In the literature, many researchers propose detailed analysis and conceptual models to explicate the complex relationship between land use and travel behaviour. Among the several land use characteristics, the structural characteristics of neighbourhoods play a relevant influence on transportation variables, as trip generation, trip length and modal split (Handy et al., 2005; Kitamura et al., 1997). In particular, these studies show that residents of neighbourhoods with higher levels of urban density, land-use mix, transit accessibility, and pedestrian friendliness (among other characteristics) drive less than residents of neighbourhoods with lower levels of these characteristics. However, the authors affirmed that, in order to provide a reliable analysis of the phenomena, socio-economic characteristics and travel attitudes of residents could be taken into account. These models have a strong spatial component because the data regarding land use, population, and transportation system have spatial references (Mazzulla and Forciniti, 2012). In addition to spatial variation, the study of travel demand shows the categorical variation of variables. Some examples are variables regarding socio-economic characteristics of population, as occupational status or education level, but also variables related to transportation system, as available modal choices and amount of trips, and the accessibility levels. For taking into account the categorical variation of variables, ordered regression models can be applied.

The main aim of this paper is to analyse the influence of the categorical variation of variables impacting on transit use. The study case is the metro network of Madrid (Spain), where, in the last few years, the realisation of new lines and stations has caused important variations in the localisation of population and economic activities, and in users' travel behaviour. An ordered probit (OP) model is proposed for evaluating how the use of metro system depends on variables related to socio-economic characteristics of population, territorial features, accessibility, and transportation system.

The rest of the paper is organised as follows. In the next section, the characteristics of the study case are illustrated. Section 3 describes the methodological approach and provides some theoretical remarks about OP models; the same section contains a discussion about the proposed model and the choice of the variables, and the main results. Finally, the paper ends with brief conclusions about the work.

2. STUDY CASE

Madrid has experienced important modifications during the past 50 years, changing from one-nuclear city to a poly-nuclear metropolis (Monzón and de la Hoz, 2009). The city was invested by an intense developing process accompanied by the dispersion of urban settlements. The urban sprawl has led to the appearance of the Madrid metropolitan area, constituted by the capital city together with a set of smaller neighbouring towns.

Analysing urban structure of Madrid municipality, more central zones are mainly areas

addressed to residential buildings, surrounded by a ring of areas addressed to economic activities and to infrastructural equipments. In the last 15-20 years, Madrid's population decreased whereas the number of inhabitants has been increasing in the surrounding towns. This could be due to many factors, and mainly because Madrid city centre is a consolidated historical area, surrounded by areas with a high potential growth and low land prices. In addition, the transit system expansion caused differential growth.

Madrid's metro system opened in 1919 with a total length of 3.48 kilometres and 8 stations. It has been in constant development ever since. The most recent extensions made the metro network 293 kilometres long by 2010 and subway became an interurban system that reached several towns in the metropolitan area. In addition, Madrid's metro system is interconnected with the light rail system and with the suburban railways servicing short distance travel to and across the city.

Data collected by the Household Travel Survey of Madrid (EDM), conducted in 2004 by the transportation company of Madrid Region (CRTM, 2006), provide a clear outline of the situation about the mobility in the metropolitan area. An amount of 14,511,397 trips was recorded, referred to a normal weekday in the period October - November of the year 2004. About 60.0% of trips are made for study or work purposes (systematic mobility). An average number of 2.60 daily trips per inhabitant was recorded. The number of trips per inhabitant increased of 20.3% compared to the results with the same type of survey conducted in 1996. In regard to the modal split, 54.7% of trips were made by transit system and 45.3% by private vehicles. The most chosen transit system is metro (37.3% of users). The entire area showed a decrease of trips made by private vehicles.

3. ORDERED PROBIT MODEL

3.1 Theoretical framework

Models for evaluating travel demand have a strong spatial component. In addition to spatial variation, the variables that characterise the interaction between land use and transportation system can be considered as categorical variables and measured with a scale consisting of a set of categories. Ordered regression models are convenient for treating the categorical variation of variables; for this reason, they have a good applicability for analysing land use-transport interaction and the accessibility to transit system.

Ordered regression models are divided into two groups: ordered probit and ordered logit, according to the statistical errors distribution. Since many years ago, several researchers have indicated that the results are similar; however, there is no consensus on which model is the best. The authors of this paper have already adopted in the past OP models for analysing service quality of airport transit services (Eboli and Mazzulla, 2009), and road safety (Cardamone et al., 2014, 2015, de Ona et al., 2014).

In the OP model there is an observed ordinal variable Y, which is, in turn, a function of another variable Y^* that is not measured (Borooah, 2001). Specifically, in the ordered model there is a continuous unmeasured latent variable Y^*, whose values determine what the observed ordinal variable Y matches. The continuous latent variable Y^* has various threshold points. The value Y_i of the observed variable depends on whether or not it crossed a particular

threshold, as showed by the following formulas (1).

$$Y_i = 1 \; if \; Y_i^* \le k_1$$

$$Y_i = 2 \; if \; k_2 \le Y_i^* \le k_1$$

$$Y_i = j \; if \; k_j \le Y_i^* \le k_{j-1} \tag{1}$$

$$Y_i = M \; if \; Y_i^* \ge k_{M-1}$$

In the population, the continuous latent variable Y^* is equal to (2):

$$Y_i^* = \sum_{k=1}^{K} \beta_k X_{ki} + \varepsilon_i = Z_i + \varepsilon_i \tag{2}$$

where ε_i is a random disturbance term normally distributed. The error term reflects the fact that the variables may not be perfectly measured, and some relevant variables may be not introduced in the equation.

By means of the OP we can estimate the expected average value of the Y_i^* (3):

$$E\left(Y_i^*\right) = Z_i = \sum_{k=1}^{K} \beta_k X_{ki} \tag{3}$$

Once we have estimated β coefficients and the (M-1) k cutoff terms, we can estimate the probability that Y will have a particular value. The formulas are the following (4):

$$P(Y = 1) = 1/\left[1 + \exp(Z_i - k_1)\right]$$

$$P(Y = 2) = 1/\left[1 + \exp(Z_i - k_2)\right] - 1/\left[1 + \exp(Z_i - k_1)\right] \tag{4}$$

$$P(Y = M) = 1 - 1/\left[1 + \exp(Z_i - k_{M-1})\right]$$

Finally, the OP model can be used to estimate the probability that the unobserved variable Y^* falls within the various threshold limits.

3.2 Proposed model

In the proposed OP model, the dependent variable Y relates to the use of metro system, according to a two-point numerical scale. The dependent variable is equal to 0 if the number of generated daily trips made by metro system is smaller than 50% of the total trips, and equal to 1 if the trips are more than 50%, and is defined in the following way:

$$Y_i = 0 \; if \; Y_i < k_1$$
$$Y_i = 1 \; if \; Y_i > k_1 \tag{5}$$

where i indicates the unit of analysis (zone of the city of Madrid) and k_1 is the threshold parameter estimated by model together with the parameters β. In order to explicate the dependent variable, a set of independent variables was defined on the basis of a theoretical reasoning linking some factors to the use of metro system. Each factor was treated making a priori hypotheses and estimating an expected relationship.

The first three factors ("Female gender", "Occupational status" and "No car ownership") relate to socio-economic characteristics of population. "Land use prevalently residential" and "Predominance of services among economic activities" are factors regarding land use

characteristics. Lastly, "Accessibility" is a factor that measures the ease to access in all the other zones, and "Availability of metro stations" refers to metro system characteristics. The factors were defined on the basis of the official zoning of the Madrid municipality.

"Female gender" relates to the number of females in total population. Usually, females, compared to males, prefer more to move using transit systems rather than private car. The factor "Worker" refers to the occupational status, and considers the amount of workers compared to the total population. Workers make more trips than unemployed inhabitants, and, generally, trips made by all transport mode increase. Car ownership is taken into account in the factor "No car ownership", which considers the households that have not a car. People without car have to move using transit systems.

The factor "Land use prevalently residential" regards the amount of areas addressed to residential settlements compared to the amount of the other land uses. The predominance of residential building in a zone could produce the necessity to move for making daily activity, as reaching own job or shopping. "Predominance of services among economic activities" refers to the amount of services (public and private) compared to other economic activities. People that live in a zone where a lot of services are localised do not need to travel towards another zone, and also trips made by metro system decrease.

The factor "Accessibility" measures accessibility of a certain zone defined as the ease to access in all the other zones. In this work, active accessibility AA_o was measured on the metro network using a gravitational model:

$$AA_o = \sum_{d=1}^{n} \left(Add_o / D_{od} \right) \tag{6}$$

where Add_d is the number of total jobs in the zone representing the trip destination, and D_{od} is the distance between the origin and the destination zones measured on the metro network. Higher level of accessibility makes metro system as more attractive.

The factor "Availability of metro stations" refers to the number of metro stations that can be reached walking less than 600 meters from home (Calvo et al., 2013). Availability of at least one metro station promotes the use of metro system.

Each factor was defined by two explanatory variables, except the factor "Accessibility", which is defined by three variables. In the case of two variables, the first variable (with value "0") corresponds to the case opposing to the factor, whereas the second variable (with value "1") corresponds to the case according to the factor.

In the case of factor "Female gender", the variable is equal to value "1" when the number of females is higher than the number of males in the considered zone, "0" otherwise. One of the variables of factor "Worker" has the value "1" if the number of workers is higher than the number of people who do not work. Considering the factor "No car ownership", the variable values "1" when the number of households that have not a car is higher than the number of households that have at least one car. The variable of the factor "Land use prevalently residential" is equal to "1" if the residential settlements are prevalent in comparison with the amount of the other land uses in the same zone. "Predominance of services among economic activities" is equal to "1" if the amount of services (public and private) is prevalent compared to other economic activities. The variable of the factor

"Availability of metro stations" is "1" when there is the availability of at least one metro station that can be reached walking less than 600 meters from home. In the case of factor "Accessibility", the first variable corresponds to the lowest level of accessibility, the second to the intermediate level, and the third to the highest level.

3.3 Results

Two OP models were elaborated in order to identify what model better explains the data. The first model contains all the variables previously specified. In order to calibrate the coefficients for each variable, the model was based on a particular reference case, corresponding to the variable of the highest level for each factor. Definitively, we have a total of 7 factors and 15 independent variables. The statistics on the goodness of fit are adequate. Based on the p-values of the Wald tests, four variables are found to be significant with p<0.1, and two variables with p<0.2; the other two variables cannot be considered as significant (table 1).

In order to improve the number of significant variables, after some attempts we decided to reduce the number of variables. In this way, the final model was elaborated considering the variables related to the factors "Female gender", "Land use prevalently residential", "Predominance of services among economic activities", and "Accessibility". The model was based on a particular reference case, corresponding to female gender, land use prevalently residential, predominance of services among economic activities, and high level of accessibility. In this case, we have a total of 4 factors and 9 independent variables. As we can observe in table 1, all the variables are significant.

Observing the sign of the estimated coefficients β, we can say that a priori hypotheses about the theoretical reasoning link between dependent and independent variables are verified. The sign of the variable V1=0 is negative; this means that if the population is mainly composed by males the trips made by metro system decrease. The sign of V7=0 is negative; therefore the number of trips generated from a zone with a low amount of areas addressed to residential building is lower compared to a zone where land use is prevalently residential. A positive sign is associated to the coefficient of the variable V9=0, therefore the number of trip made by metro is higher from a zone with a lower presence of services compared with a zone with predominance of services among economic activities. As expected, the variables V11=1 and V12=2 have negative sign because lower levels of accessibility on metro network negatively impact on the number of trips made by metro.

By observing the estimated probabilities we can state that the probabilities of the reference case are about 84% for daily trips made using metro system and about 16% for daily trips made using other motorised transportation means. The estimated probabilities related to variables with negative coefficients show that the probability to travel by metro system decreases compared to the variable considered as reference case. Instead, in the case of the variable V9=0, whose coefficient has positive sign, the probability to made trips by metro system increases compared to the probabilities calculated for the variable considered as reference case. The variable that mainly impacts on the choice of metro system is V1=0, related to the factor "Female gender". When the travellers are males, the probability to have

trips made by metro decreases to about 45%, and, at the same time, the probability to use other motorized means increases to about 55%. The influence of the factor "Accessibility" is also relevant on the drop of the probability to made trips by metro.

Factor	Variable	Model 1			Model 2				
		β	Wald	p-value	β	Wald	p-value	Estimated probability	
								0	1
Reference case								0.1641	0.8359
Female gender									
	[V1=0]	-1.120	3.011	0.083	-1.067	3.118	0.077	0.5486	0.4514
	[V2=1]	0	.	.	0	.	.	0.2389	0.7611
Worker									
	[V3=0]	-0.292	0.984	0.321					
	[V4=1]	0	.	.					
No car ownership									
	[V5=0]	-0.290	0.909	0.340					
	[V6=1]	0	.	.					
Land use prevalently residential									
	[V7=0]	-0.514	1.959	0.162	-0.669	3.827	0.050	0.3379	0.6621
	[V8=1]	0	.	.	0	.	.	0.0936	0.9064
Predominance of services among economic activities									
	[V9=0]	0.628	3.595	0.058	0.701	4.986	0.026	0.2508	0.7492
	[V10=1]	0	.	.	0	.	.	0.2630	0.7370
Accessibility									
	[V11=1]	-1.181	3.674	0.055	-1.317	5.051	0.025	0.3872	0.6128
	[V12=2]	-1.331	10.696	0.001	-1.462	13.934	0.000	0.3709	0.6291
	[V13=3]	0	.	.	0	.	.	0.0607	0.9393
Availability of metro stations									
	[V14=0]	-0.408	1.768	0.184					
	[V15=1]	0	.	.					
Number of observations		128			128				
k_1 (threshold)		-2.319			-1.853				
ρ^2 (Cox and Snell)		0.240			0.216				
ρ^2 (Nagelkerke)		0.352			0.317				
ρ^2 (McFadden)		0.240			0.213				
log likelihood		-32.841			-12.790				

Table 1 – Ordered probit model results

4. CONCLUSION

In this paper, an ordered probit (OP) model was proposed for exploring how factors related to socio-economic characteristics of population, territorial features, accessibility, and transportation system, that impact on transit use. The case study was the metro network of Madrid (Spain). The explanatory variables were selected on the basis of the analysis of the theoretical reasoning linking some factors to the use of metro system. At first, seven factors were selected, but after some attempts, we proposed a final model considering four factors, that are "Female gender", "Land use prevalently residential", "Predominance of services among economic activities", and "Accessibility".

The results showed that the variable that mainly impacts on the choice of metro system is related to the factor "Female gender". When the travellers are males, the probability shows a clear inclination to travel using other motorised transportation modes rather than metro system. Relevant results are also the influence of the factor "Accessibility" on the drop of the probability to made trips by metro. Factors regarding land use characteristics, as "Land

use prevalently residential" and "Predominance of services among economic activities", showed also a certain influence on the probability to use the metro system. The other socio-economic characteristics of the population ("Occupational status" and "No car ownership") seem to be uninfluential, as well as the factor "Availability of metro stations".

Definitively, the results confirmed the hypothesis we made about the influence of the categorical variation of variables affecting the use of metro system. In this context, by operating on accessibility and supplied transit services, travel demand can be shift from the individual means of transportation towards transit systems.

REFERENCES

BOROOAH, V.K. Logit and Probit: Ordered and Multinomial Models. Sage University Papers Series on Quantitative Applications in Social Sciences, Serie no. 07-138. Thousand Oaks, CA: Sage, 2001.

CALVO, F.J., DE OÑA, J., and ARÁN, F. (2013). Impact of the Madrid subway on population settlement and land use. *Land Use Policy* 31, pp. 627-639.

CARDAMONE, A.S., EBOLI, L., FORCINITI, C., and MAZZULLA, G. (2014), Willingness to use mobile application for smartphone for improving road safety, *International Journal of Injury Control and Safety Promotion*. DOI:10.1080/17457300.2014.945463.

CARDAMONE, A.S., EBOLI, L., FORCINITI, C., and MAZZULLA, G. (2015), How usual behaviour can affect perceived drivers' psychological state while driving, *Transport*. DOI: 10.3846/16484142.2015.1059885.

CRTM (2006). *Encuesta domiciliaria de movilidad en día laborable de 2004 en la Comunidad de Madrid. Resumen*. Consorcio Regional de Transportes de Madrid.

DE OÑA, J., DE OÑA, R., EBOLI, L., FORCINITI, C., and MAZZULLA G. (2014), Key factors affecting drivers' perception of accident risk. *Accident Analysis and Prevention* 73 (1), pp. 225-235.

EBOLI, L., and MAZZULLA, G. (2009). An ordinal logistic regression model for analysing airport passenger satisfaction. *Euromed Journal of Business* 4(1), pp. 40-57.

HANDY, S., CAO, X., and MOKHTARIAN, P. (2005). Correlation or causality between the built environment and travel behavior? Evidence from Northern California. *Transportation Research Part D* 10, pp. 427-444.

KITAMURA, R., MOKHTARIAN, P., and LAIDET, L. (1997). A micro-analysis of land use and travel in five neighborhoods in the San Francisco Bay Area. *Transportation* 24, pp. 125-158.

MAZZULLA, G., and FORCINITI, C. (2012). Spatial association techniques for analysing trip distribution in an urban area. *European Transport Research Review* 4, pp. 217-233.

MONZÓN, A., and DE LA HOZ, D. (2009). Efectos sobre la movilidad de la dinámica territorial de Madrid. *Urban* 14, pp. 58-71.

NOCERA, S., MAINO, F., and CAVALLARO, F. (2012). A heuristic method for determining CO_2 efficiency in transportation planning. *European Transport Research Review. An Open Access Journal* 4(2), pp. 91-106.