Examining the Impact of Household Interactions when Modelling Travel Duration

Elke Moons
Transportation Research Institute, Hasselt University
Science Park 5/6
B-3590 Diepenbeek, Belgium
E-mail: elke.moons@uhasselt.be

Geert Wets
Transportation Research Institute, Hasselt University
Science Park 5/6
B-3590 Diepenbeek, Belgium
E-mail: geert.wets@uhasselt.be

1. Introduction

From an international research perspective, an activity-based view on transportation has become standard today. Instead of modelling travel demand separately, the major idea behind these activity-based models is that travel demand is derived from the activities that individuals and households need or wish to perform (Jones, *et al.*, 1983). In a modern society, mobility is considered to be vital for a human's development: it is not only regarded as one of the factors behind economic growth, but also as a social need that offers people the opportunity for self-fulfillment and relaxation (Ministry of Transport and Public Works, 2004). In the light of activity-based models, this seems a logical consideration. Travel patterns are regarded as the manifestation of the implementation of activity programs over time and space. In turn, activity patterns emerge as the interplay between the institutional context, the urban/physical environment, the transportation system and individuals' and households' needs to realize particular goals in life and to pursue activities (Ben-Akiva and Bowman, 1998). The aim of these models is to predict which activities will be conducted where, when, for how long, with whom and with which transport mode and as a consequence, they require a huge amount of data to do so.

Travel diaries are currently one of the most important ways of obtaining the critical information needed for transportation planning and policy development. These diaries are used to collect current information about the socio-demographic, economic and trip-making characteristics of individuals and households. Not only travel characteristics (transport mode, duration of travel,. . .) are important, also household aspects (e.g. with whom the activity is conducted, number of children in a household, household income) and individuals features (age, gender, etc.) need to be collected. This clearly shows why household travel surveys, combined with individual surveys, continue to be an essential component of transport planning and modelling efforts. Activity-diaries mainly form the basis of an activity-based survey, and next to these individual questionnaires, also a household survey needs to be filled out. One should therefore also acknowledge that both levels can and will (Jovicic, 2001) play an important role in trying to predict each of the responses stated above. Very often, though, only household variables are taken into account in an attempt to model a response at an individual level, without actually accounting for the clustering that is present. Another option that is pursued sometimes is to account for this information at different levels by posing constraints on the underlying model. However, there is another option: it is possible to incorporate the correlation between the household and the individual level directly by modelling it by means of mixed models. Section 2 discusses the theoretical background of the models, while Section 3 describes the application of these models to a very important response in the Flemish household travel survey, the time spent on travelling. Finally, Section 4 gives a general conclusion and some avenues for future research.

2. Methodology

The mixed model methodology has been developed within the discipline of animal genetics and breeding (because of possible correlations between individual animals and their herd), but from there, it has spread to many other disciplines, such as medicine, sociology, etc.. It can be used in any context in which observations are correlated with each other, e.g. because they are correlated in time or because they are spatially correlated (Aerts, *et al.*, 2002). In this paper, we suggest to apply the mixed model approach to the transportation context, because of the clear and present correlations at different levels within a household. An advantage of the Generalized Linear Mixed Model (GLMM) methodology (Verbeke and Molenberghs, 2000), which will be applied in Section 3, is that they can account for the problem of missing data as well, whereas some of the very well known techniques (such as multivariate models) might not be valid in the presence of missingness.

The methodology can shortly be described as follows. If data is gathered at different levels, one may assume that there is a hierarchical structure between them (Goldstein, 1995). In this paper, we will show a classical example of a two-level hierarchical structure. Level 1 is the level of the smallest unit (individual) whereas the second level denotes the clusters of the units (household). The main idea of a GLMM is to examine the behaviour of the level 1 outcome as a function of predictors that behave both on level 1 and on level 2.

The outcome of individual i in household j model can in general be written as:

$$\begin{aligned} Y_{ij} &= \beta_{0j} + \sum_{k=1}^{K} \beta_{kj} X_{kij} + \varepsilon_{ij} \\ \text{with } \beta_{kj} &= \alpha_{k0} + \sum_{p=1}^{P} \alpha_{kp} Z_{pj} + u_{kj}, \text{ for } k = 1, \ldots, K; \end{aligned}$$

$$\varepsilon_{ij} \sim N(0, \sigma^2) \quad \text{and} \quad \begin{pmatrix} \mathrm{u}_{0j} \\ \vdots \\ \mathrm{u}_{Kj} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \cdots & \sigma_{0K} \\ \vdots & \ddots & \vdots \\ \sigma_{K0} & \cdots & \sigma_K^2 \end{pmatrix}\right].$$

The X-variables refer to predictors at the first level (individual), whereas the Z's are explanatory variables at the second level (household). If one only wants to account for predictors at the second level, then all other β-parameters, except β_{0j} can be set equal to zero. Similarly, only accounting for explanatory variables at the first level can be carried out by setting all α-parameters, except α_{k0} equal to zero. The index p behaves at household level, while index k is associated with the individual level in this paper. Some variables might play a role at both levels, however, in this paper, we restricted ourselves to a strict distinction between both levels.

3. Results

3.1 The data

The Flemish travel survey for the year 2000 (Zwerts and Nuyts, 2004) will be used for the analyses presented here. People were asked to write down for some consecutive days which activities they conducted, where, when, with whom, for how long and which transport mode was used to arrive at the location of the activity. Above this information, some general household information was gathered as well, such as household composition, socio-economic status of the household, availability of transport modes, etc. Trips of all road users (car drivers, car passengers, pedestrians, bike and motorbike riders and public transport users) were registered for the period January 2000 - January 2001. It is based on a random sample of 2 823 households, including 7 638 people who were more than 6 years old. In total 21 031 trips were registered. This survey had a response rate of 32%.

3.2 Analyses

In this subsection, the models described in Section 3 will be applied to the response variable, travel duration. Previous results have shown already that up to 30% of the variation in travel time can be attributed to the fact that people live in households (Moons and Wets, 2007). Since it seems logical that the impact of clustering can be different for various travel purposes, three travel goals are investigated in this paper: commuting trips (work and school related), shopping trips and leisure trips.

The explanatory variables that are considered at household level are: the total household income (HHinc), the number of children younger than 6 years (nrkid6), the number of people in the household (nrpHH), the number of cars (nrcar) and the number of bikes (nrbike) in the household. At individual level the age of the person (age), the gender (sex) and their diploma (dipl) might play an important role. Table 1 will indicate the significant variables at different levels within the hierarchical structure. Significance is determined at 5% significance level. Models will be fitted including effects separately at household level and at individual level, and at both levels at the same time.

Table 1: Significant variables for different models

Purpose	Level	HHinc	nrkid6	nrpHH	nrcar	nrbike	age	sex	dipl
Commuting	Level 1						X	X	
	Level 2				X	X			
	Level 1 & 2				X	X	X	X	
Shopping	Level 1						X		
	Level 2			X					
	Level 1 & 2						X		
Leisure	Level 1						X		
	Level 2			X		X			
	Level 1 & 2						X		

It may seem strange that when variables are included at both levels that only predictors of one level seem to be significant. For each model a separate backward selection procedure was carried out and the final models ended up having the above variables.

The intra-class correlation for a model without predictors determines the portion of the total variance that occurs between households (Singer, 1998). For commuting trips this equals 9.62%, for shopping trips 53.92% and for leisure trips 57.74%. This learns us that especially for shopping and leisure trips, there is quite a bit of clustering in travel time within the households. Hence, carrying out an ordinary least squares regression analysis on these data, would likely yield misleading results. It seems only logical that this is not particularly true for commuting trips.

If we take a look at the individual level, we can determine how much of the explainable variation within the households can be attributed to the predictors at individual level. For commuting trips, this is 22.49%, for shopping trips, age seems to explain as little as 6.36% of the within-household variation, whereas for leisure trips it explains 27.90%.

At household level, we can determine a similar figure, that shows how much of the household-to-household can be explained by the explanatory variables at household level. For commuting trips 15.65% can be explained by the number of cars and the number of bikes. 2.01% of the explainable variation in shopping trips is attributable to the number of people in the household and for leisure-related trips the household size and the number of bikes explain 6.81% of the household-to-household variation. The intra-class correlation determines whether there is still any variation in households remaining to be explained. For commuting trips the intra-class correlation yields 0.08, showing that the travel time among individuals within households after controlling for the number of cars and the number of bikes is not very similar. For shopping trips, the intra-class correlation equals 0.54 and for leisure trips it is 0.56. After controlling for household-related variables, the travel time for both shopping and leisure trips is quite similar among individuals within households.

4. Conclusions and further research

Modelling travel behaviour has always been a major area of concern in transportation research, but it's importance becomes only larger today. In order to lead an efficient policy, governments require reliable predictions of travel behaviour. Most of the data sets collected in transportation research for the modelling of travel behaviour show a multilevel or hierarchical structure. These structures are common in practice and it can be argued that they are the norm, rather than the exception. However, the literature that discusses these hierarchical models in the transportation context is rather limited. Judging by the fact that for trips other than commuting trips, there is quite some clustering present in the travel time within households, this is not something that may be overlooked. The results in this paper show that even up to 56% of the variation in travel time can be attributed due to the fact that people live in households. Judging by these results, the importance of correlation structures cannot be ignored anymore as it is too often done today. Future research will focus on other response variables, such as travel distance.

REFERENCES

Aerts, M., Geys, H., Molenberghs, G. and Ryan, L. (2002) *Topics in Modelling of Clustered Data.* Chapman & Hall, London.

Ben-Akiva, M.E. and Bowman, J.L. (1998) Integration of an activity-based model system and a residential location model. *Urban Studies*, **35(7)**, 1231–1253.

Goldstein, H. (1995) *Multilevel Statistical Models - 2nd edition.* Edward Arnold, London.

Jones, P.M., Dix, M.C., Clarke, M.I. and Heggie, I.G. (1983) *Understanding Travel Behavior.* Gower, Aldershot.

Jovicic, G. (2001) *Activity Based Travel Demand Modelling.* Danmarks TransportForskning.

Ministry of Transport and Public Works (2004) *Nota Mobiliteit (in Dutch).* Ministerie van Verkeer en Waterstaat, Den Haag.

Moons, E. and Wets, G. (2007) Modelling household interactions: A mixed model approach. Paper to be presented at the 11^{th} World Conference on Transportation Research.

Singer, J.D. (1998) Using SAS proc mixed to fit multilevel models, hierarchical models and individual growth models. *Journal of Educational and Behavioural Statistics*, **24(4)**, 323–355.

Verbeke, G. and Molenberghs, G. (2000) *Linear Mixed Models for Longitudinal Data.* Springer Series in Statistics.

Zwerts, E. and Nuyts, E. (2004) *Onderzoek verplaatsingsgedrag Vlaanderen 2 (in Dutch).* Publication D/2004/3241/016. Provinciale Hogeschool Limburg, Diepenbeek, Belgium.