Modeling Household Interactions: A Mixed Model Approach

Elke Moons*
Post-doctoral researcher
Hasselt University – Campus Diepenbeek
Transportation Research Institute (IMOB)
Wetenschapspark 5, bus 6
3590 Diepenbeek
Belgium
Tel.: +32 – 11 – 26.91.26
Fax: +32 – 11 – 26.91.99
E-mail: elke.moons@uhasselt.be

Geert Wets
Full professor
Hasselt University – Campus Diepenbeek
Transportation Research Institute (IMOB)
Wetenschapspark 5, bus 6
3590 Diepenbeek
Belgium
Tel.: +32 – 11 – 26.91.38
Fax: +32 – 11 – 26.91.99
E-mail: geert.wets@uhasselt.be

Word count: 5522 words
Revision date: May 1, 2007

* corresponding author

ABSTRACT
From an international research perspective, an activity-based view on transportation has become standard today. The aim of activity-based models is to predict which activities will be conducted where, when, for how long, with whom and with which transport mode. The required information for these models is gathered by means of the activity-diary surveys. These diaries provide information at the individual level, but next to that, also a household survey needs to be filled out. One should therefore acknowledge that both levels can 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. This paper gives a better solution by incorporating the correlation between the household and the individual level directly by means of mixed models. The mixed model strategy is applied to two important responses of the Flemish travel survey of 2000, i.e. daily travel time and daily travel distance. The employment status, gender and age are shown to be important individual characteristics, while the number of children and the household income are significant household variables. Moreover, the results show that up to 30% of the variation in travel time and/or distance can be attributed due to the fact that people live in households, so this may not be ignored any longer.

1. Introduction

Modeling travel behavior has always been a major area of concern in transportation research. Since 1950, due to the rapid increase in car ownership and car use in Western Europe and in the US, several transportation planners attempted to model people's choice of transport mode, route choice and destination. The resulting trip-based models were employed to predict travel demand on the long run and to support investment decisions in new road infrastructure that originated from this increased level of car use. At the time, travel was assumed to be the result of four subsequent decisions that were modeled separately: trip generation, trip production, mode choice and route choice (McNally, 2000). However, one acknowledged that these initial models clearly had some drawbacks (Jovicic, 2001), like e.g. the focus on individual trips, where the interrelationships (spatial, temporal, intra-household) between trips and its characteristics are ignored.

The ensuing tour-based models that were developed in the seventies and eighties accounted for this problem, but still limited insight was offered into the relationship between travel and non-travel aspects, since travel has an isolated existence in these models and the question why people undertake trips is completely neglected. This is where activity-based models come into play, setting the standard for the last decade of modeling travel demand. The major idea behind such 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 turn, activity patterns emerge as the interplay between the institutional context, the urban/physical environment, the transportation system and individual's and household's needs to realize particular goals in life and to pursue activities (Ben-Akiva and Bowman, 1998). Activity-based models aim to predict these activity patterns of each individual, and require a huge amount of data to do so.

Travel surveys are currently one of the most important ways of obtaining the critical information needed for transportation planning and policy development. These surveys are used to collect current information about the demographic, socio-economic, and trip-making characteristics of individuals and households as well as to further our understanding on travel in relation to the choice, location, and scheduling of daily activities. 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 individual 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 modeling 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 modeling it by means of mixed models.

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 (Goulias, 2002).

In the past, too few attention has been posed to this kind of problems. Moreover, very often in activity-based surveys, one also has to account for the problem of missing data for one or more persons in a household. Some of the very well known techniques (such as multivariate models, e.g.) might not be valid in case of missing data. Generalized Linear Mixed Models (GLMM) (Verbeke and Molenberghs, 2000; Molenberghs and Verbeke, 2005) can take care of this. In this paper, we will present the results of GLMM applied to some important responses of the Flemish travel survey of 2000, such as the time that has been spent on traveling during a day, and the daily traveled distance. A sequence of models will be fit: at first a regular ANOVA model, next a model where only household characteristics play a role, then a similar model but only using individual characteristics, and finally a mixed model that combines both. The results will be compared to those of the often used multivariate models and the differences between both approaches will be discussed.

In Section 2, the data that are used for the analyses will be presented. Section 3 discusses the theoretical background of the models, while Section 4 describes the application of these models to the Flemish travel survey data. Finally, Section 5 gives a general conclusion and some avenues for future research.

2. The Data

In a typical household travel survey, a sample of the population is asked to record their activity patterns over a given time period. This information is combined with socio-demographic information about the sample to develop relationships between individual/household characteristics and their observed travel patterns. More precise, people are 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. Traditionally, travel data on households and individuals are collected about every five years in Flanders (Dutch speaking part of Belgium) on about two thousand five hundred households that have to report their travel behavior. Each person in the household which is older than six years is asked to fill out the travel diary. This household sample is different at each wave of surveying.

The Flemish travel survey for the year 2000 (Zwerts and Nuyts, 2004) will be used for the analyses presented in Section 4. 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. Multilevel Models

3.1 The Unconditional Means Model

At first, the variation in the response variables across households is examined by means of an unconditional means or a one-way random effects ANOVA model. The ANOVA-way of writing down this model expresses the outcome, Y_{ij}, as a linear combination of the grand mean (μ), household deviations from that mean (α_j) and a random error (ε_{ij}) associated with the i-th individual in household j:

$$Y_{ij} = \mu + \alpha_j + \varepsilon_{ij}$$

$$\text{with } \alpha_j \sim_{\text{iid}} N(0,\sigma_H^2) \text{ and } \varepsilon_{ij} \sim N(0,\sigma^2).$$

This model will now be re-parameterized to the general multilevel notation, because this notation can be generalized more easily to the more complex models.

It expresses the level 1 outcome by means of set of linked models: one a the individual level and one at the household level. At level 1, the outcome can be denoted as the sum of an intercept for the individuals household (β_{0j}) and a random error (ε_{ij}) associated with the i-th individual in household j:

$$\text{Level 1: } Y_{ij} = \beta_{0j} + \varepsilon_{ij} \qquad \text{with } \varepsilon_{ij} \sim N(0,\sigma^2).$$

At the second level, the households intercept is expressed as a sum of the overall mean (μ) and a series of random deviations from that mean (α_j):

$$\text{Level 2: } \beta_{0j} = \mu + \alpha_j \qquad \text{with } \alpha_j \sim_{\text{iid}}. N(0,\sigma_H^2)$$

Substituting the level 2 model in the level 1 equation yields the multilevel model:

$$Y_{ij} = \mu + \alpha_j + \varepsilon_{ij}$$

$$\text{with } \alpha_j \sim_{\text{iid}} N(0,\sigma_H^2) \text{ and } \varepsilon_{ij} \sim N(0,\sigma^2).$$

Note that we also assume that α_j and ε_{ij} are independent of one another. One can notice that there is a direct equivalence between the one-way random effects ANOVA notation and the multilevel notation.

This model can be partitioned into two separate parts: a fixed part that contains the single effect μ (the overall intercept) and a random part that contains two random effects (the intercepts α_j and the within-household residuals ε_{ij}). The fixed effect μ will learn us something about the average outcome in the population, the parameter of the first random effect σ_H^2 tells us about the variability in the household means, while σ^2 tells us something about the variability of the outcome within the households.

This particular unconditional means model postulates that the variance-covariance structure takes a special form, i.e. that of *compound symmetry*. This means that the variance for each individual is assumed to be $\sigma_H^2 + \sigma^2$, the covariance of the outcome for two individuals of the same household equals σ_H^2, while the covariance of the outcomes for two individuals belonging to a different household is 0.

3.2 Effects at Level 2 (Household)

The unconditional means model provides a baseline against which more complex models can be compared. At first, predictors on level 2 will be included. Several explanatory variables behave at this household level: the distance between the nearest bus, tram, metro station or train station and the home location, the number of cars and bikes in a household, the number of children and the household income. This first conditional model can now be written as:

$$Y_{ij} = \beta_{0j} + \varepsilon_{ij}$$

$$\text{with } \beta_{0j} = \varsigma_{00} + \sum_{p=1}^{P} \varsigma_{0p} X_{pj} + u_{0j}.$$

All P explanatory variables behave at the household level.

Substituting this two level equation into the level 1 equation gives:

$$Y_{ij} = \varsigma_{00} + \sum_{p=1}^{P} \varsigma_{0p} X_{pj} + u_{0j} + \varepsilon_{ij}$$

with $u_{0j} \sim N(0, \sigma_H^2)$ and $\varepsilon_{ij} \sim N(0, \sigma^2)$.

3.3 Effects at Level 1(Individual)

Including *K* variables at level 1 can be carried out as follows:

$$Y_{ij} = \beta_{0j} + \sum_{k=1}^{K} \beta_{kj} X_{kij} + \varepsilon_{ij}$$

with $\beta_{kj} = \varsigma_{k0} + u_{kj}$, for *k=0,…,K*;

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

As can be observed, by including variables at the fixed effect level (ς_{k0}), additional random effects (u_{kj}) are included too. In this way, we stipulate that the outcome variable does not only depend on the explanatory variables at individual level, but also that the relationship between the explanatory variables can vary across households.

3.4 Effects at Level 1 and 2 (Individual and Household)

Combining the two previous model leads us to a model that contains variables that behave at individual level and next to that also variables that behave at household level. This combination can be written down as follows:

$$Y_{ij} = \beta_{0j} + \sum_{k=1}^{K} \beta_{kj} X_{kij} + \varepsilon_{ij}$$

with $\beta_{kj} = \varsigma_{k0} + \sum_{p=1}^{P} \varsigma_{kp} X_{pj} + u_{kj}$, for *k=0,…,K*;

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

The index *p* behaves at household level, while the index *k* behaves at individual level. Some variables might play a role at both levels, however, in this paper, we restricted ourselves to a strict distinction between both level. Next to this, to avoid too complex models, only main effects of the variables are incorporated, no interaction effects (Kutner, *et al.*, 2004).

4. Analyses

In this Section, the four types of models described in the previous Section will be applied to travel time and travel distance. It seems logical to believe that the inherent relation between individuals who live together in a household has a certain impact on these two outcome variables. However, it requires some effort to investigate whether this is really true. If this hypothesis proves to be true, we can still pose the question to what extent this is the same for both outcome variables.

Individuals and households provide a classical example of two-level hierarchical models (Goldstein, 1995). The main idea is to examine the behavior of the level 1 outcome as a function of predictors that behave both on level 1 (individual) and on level 2 (household). For this purpose, a series of models with increasing complexity is tested. The first model is an unconditional means or a one-way random effects ANOVA model. This shows how much of the variation in the data can be captured by allowing solely a separate intercept for each household.

In the next models, explanatory variables come in at different levels. The second model includes predictors at the household (level 2) level, while the third model takes explanatory variables at the individual level into account. The fourth model, finally, allows for effects at both levels.

<u>4.1 The Unconditional Means Model</u>

a. Travel Time

The fixed effect parameter ($\hat{\mu}$) equals 3.9202 (S.E. = 0.0150), which would indicate that the average travel time on a household basis is about 50.41 min daily for this sample of households. Note, that this is not the same as the average individual travel time. In general, this is somewhat low. The time per day that is spent on traveling is a constant over the years, and it equals about 1.1 hour (see Schäfer, 2000 and Schäfer and Victor, 2000). Some general goodness-of-fit measures on this model are provided as well. Minus twice the log-likelihood (ll) of this model equals 12530.7, and the AIC criterion yields 12534.7. The estimates of the random effects in the model are provided in Table 1.

<INSERT TABLE 1 ABOUT HERE>

These estimates suggest that households differ a little in their average (log) travel time and that the variation among individuals within households is more than twice as large as the variation between households. Another way of reflecting on these sources of variation in (log) travel distance is by estimating the intra-class correlation (ρ) (Singer, 1998). This figure will teach us what portion of the total variance occurs between households. It is determined by

$$\hat{\rho} = \frac{\hat{\sigma}_H^2}{\hat{\sigma}_H^2 + \hat{\sigma}^2} = \frac{0.2339}{0.2339 + 0.5398} = 0.3023 .$$

Hence, there is quite a bit of clustering of (log) travel time within households. As a consequence, an ordinary least squares (OLS) regression analysis of these data would likely yield misleading results.

b. Travel Distance

Fitting also an unconditional mean model, but now on the second response variable, the travel distance, learns us that the average travel distance on a household basis equals about 20.3 km per day ($\hat{\mu}$ = 3.0117). Once again this is the average for this particular sample of households, not on an individual basis. Minus twice the log-likelihood of this model is now equal to 15612.2, while the AIC yields 15616.2. The estimates for the random effects in the model are provided in Table 2.

<INSERT TABLE 2 ABOUT HERE>

The households seem to differ more in their average (log) travel distance when compared to the travel time. This is in correspondence to the findings of Schäfer (2000) and Schäfer and Victor (2000), who state that travel time is a constant in space as well as in time. Due to improved (and still improving) technologies people tend to travel further, but the travel time per day stays invariant. The intra-class correlation teaches us that 27.77% of the total variance occurs between households.

4.2 Effects at Level 2 (Household)

a. Travel Time

Apparently, the number of children and the household income are the only significant variables at the household level. When comparing both models based on their log-likelihood and Akaike's criterium, it can clearly be observed that this latter model performs a little worse than the unconditional model (-2 x log-likelihood = 12533.5 and AIC = 12537.5). The fixed effect parameters are given in Table 3.

<INSERT TABLE 3 ABOUT HERE>

The term for the intercept estimates the household mean when the remaining predictors are zero. This mean travel time at household level equals 59.11 minutes per day for a household with a monthly income above 5000 € without children. For households in a lower income category and for households with more children, the household's mean travel time decreases accordingly.

The random effect parameters are provided in Table 4.

<INSERT TABLE 4 ABOUT HERE>

Compared to the previous random effects ANOVA model, these parameters now have different meanings. In the previous model, there were no explanatory variables, so the components were unconditional.

Having added two predictors, they have become conditional components now. Notice that the random effect for the variance within the household (σ^2) has remained virtually unchanged, while the variance representing the variation between the households has diminished somewhat (from 0.2339 to 0.2317). This tells us that the predictors explain part of the household-to-household variation in mean (log) travel time. To be exact, one can say that $\frac{0.2339-0.2317}{0.2339}=0.01$ or 1% of the explainable variation in household travel time can be explained by the predictors ‘number of children’ and ‘household income’. Now the question can be posed on whether there is still any variation in household means remaining to be explained. This can be tested by looking at the *residual* intra-class correlation, i.e. the intra-class correlation among households of comparable number of children and household income. This intra-class correlation equals $\frac{0.2317}{0.2317+0.5393}=0.30$. This partial correlation shows the similarity in travel time among individuals within households after controlling for the effects of household income and the number of children.

b. Travel Distance

The number of bicycles in the household and the distance between home and the nearest train station appear not to be significant at a 5% level. The parameters for the other fixed effects can be found in Table 5. Minus twice the log-likelihood (-2*ll = 15484.9) and the AIC (equal to 15488.9) show clearly better results for this model when compared to the random effects ANOVA model.

<INSERT TABLE 5 ABOUT HERE>

The mean travel distance at household level for a household with a monthly income above 5000 €, no kids, no car and a distance between home and the nearest bus, tram or metro station of more than 5km equals about 21km. In general, for households with a lower income, households with children or with a bus, tram or metro station close to home, the mean travel distance decreases accordingly (if all other factors remain constant). The opposite occurs for households with 1 or more cars, their mean travel distance increases.

The covariance parameter estimates tell us something about the random effects. Parameters are provided in Table 6.

<INSERT TABLE 6 ABOUT HERE>

Again there is not so much difference between the random effects for the variance within the households for this model and the unconditional model. The explanatory variables are able to explain about 14% of the explainable household-to-household variation in mean (log) travel distance.

The intra-class correlation shows that the (log) travel distance among individuals after controlling for the explanatory variables at household level is similar for about 25%. An OLS regression does not account for this correlation, and hence it would provide misleading results.

Now, we want to investigate what the effect is of including predictors at the individual level to model. At first, this will be carried out by including only these level 1 variables, while excluding the level 2 predictors. In the final subsection, explanatory variables at both levels will be taken into account.

4.3 Effects at Level 1 (Individual)

a. Travel Time

Gender, age and employment status appear to be significant at the individual level in order to predict travel time. Goodness-of-fit measures (log-likelihood and Akaike's criterion) are inclined to prefer this model above the previous two (-2 x log-likelihood = 12322.6 and AIC = 12366.6). Let us first take a closer look at the fixed effects. The parameter estimates can be found in Table 7.

<INSERT TABLE 7 ABOUT HERE>

The estimate for ς_{00} indicates that the estimated average (log) travel time, controlling for age, gender and employment status, equals 3.523. The estimate for ς_{10} indicates that the estimated average slope representing the relationship between age and (log) travel time is equal to 0.0043. The standard errors for all fixed effects are very small, resulting in large t-statistics and low P-values. This gives us evidence to conclude that there exists a significant relationship between the age, gender and employment status of the individual and his/her travel time.

Again, the question can be posed: how much of the within-household variance in travel time is explained by the individual's age, gender and employment status? Just as we did before, in the previous models, the estimates for σ^2 for the unconditional and the conditional model will be compared. Looking back at section 4.1, we find an unconditional estimate of 0.5398, the conditional estimate for this model yields 0.4617. Inclusion of the predictors at individual level has therefore explained $\frac{0.5398-0.4617}{0.5398}=0.1447$ or 14.47% of the explainable variation within households. Comparatively speaking, the predictors at individual level explain the within-household variation in individual travel time much more than the predictors at household level explain the variation in household level travel time.

b. Travel Distance

Exactly the same explanatory variables as for travel time appeared to be important for predicting travel distance. The log-likelihood and the AIC criterion, however, seem to favor the model with predictors at household level. Minus twice the log-likelihood is here equal to 15278.0, while the AIC criterion yields 15322.0. Therefore, we will only give a concise description of this model, and in the next section we will go more into detail in the last model that combines predictors at household and individual level. The parameter estimates of the fixed effects for this conditional model can be found in Table 8.

<INSERT TABLE 8 ABOUT HERE>

The estimated (log) travel distance, controlling for age, gender and employment status equals 2.4140, meaning that on average, people travel just above 11km a day. Males travel about 1.3 times as far as females comparable in age and employment status, while employed people travel about 1.6 times as far as their unemployed counterparts. Once again, the standard errors for all fixed effects are very small, resulting in low P-values and providing enough evidence to conclude that there exists a significant relationship between the age, gender and employment status of the individual and his/her travel distance.

Let us take a closer look at how much of the within-household variance in travel distance is explained now by the individual's age, gender and employment status. The unconditional estimate of σ^2 equals 1.2314, whereas the value of this conditional model is equal to 0.8884. This means that including explanatory variables at the individual level can explain $\frac{1.2314-0.8884}{1.2314}=0.2785$ or 27.85% of the explainable variation within households. Thus, once again, the individual level predictors explain the within-household variation in individual travel distance much more than the predictors at household level explain the variation in household level travel distance.

4.4 Effects at Level 1 and Level 2 (Individual and Household)

Having separately specified models with either level 1 or level 2 predictors, we will now consider models which contain variables of both types. To achieve parallelism with Bryk and Raudenbush (1992), we will restrict ourselves to the variables that were found to be significant in the previous two steps.

a. Travel Time

As stated above, we first fitted the model with fixed effects for the number of children, the household income, age, gender and employment status and as random effects next to a random intercept also random slopes for age, gender and employment status. If we then start by interpreting the fixed effects, it immediately shows that the variable describing the household income has become insignificant in the presence of the other predictors at individual level. We can even simplify the initial model some more if we take a look at the random effects at the same time. The variance components for gender and employment status became insignificant. To ensure this, an approximate test of the null hypothesis that the change in log-likelihood is zero between the model that has gender and employment status as random effects and the model that leaves out both random effects, was carried out and compared to a chi-square distribution on 7 degrees of freedom (corresponding to the seven additional parameters). The result of this approximate test confirmed our findings. However, since gender and employment status apparently were very significant variables, we tried again to add them, but as fixed effects now, and the AIC-value dropped dramatically. Therefore, we concluded to incorporate both variables in the final model, so it can be written as:

$$Y_{ij} = \beta_{0j} + \beta_{1j} age_{ij} + \varepsilon_{ij}$$

$$\text{with} \quad \beta_{0j} = \varsigma_{00} + \varsigma_{01} nchild_j + \varsigma_{02} gender_{ij} + \varsigma_{03} employment_{ij} + u_{0j},$$

$$\beta_{1j} = \varsigma_{10} + \varsigma_{11} nchild_j + \varsigma_{12} gender_{ij} + u_{1j},$$

$$\varepsilon_{ij} \sim N(0, \sigma^2) \text{ and } \begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{10} & \sigma_1^2 \end{pmatrix}\right].$$

Minus twice the log-likelihood of the above model equals 12262.3 and the AIC value is 12270.3, indicating that this final model performs best.
The parameter estimates of the significant fixed effects are listed in Table 9.

<INSERT TABLE 9 ABOUT HERE>

Since gender and employment status are dummy variables, it might be helpful for interpretation purposes to rewrite some of the fitted models:

Male, employed:	Travel time = 4.0257 + 0.0001 Age - 0.2223 Nchild + 0.0070 Nchild*Age
Female, employed:	Travel time = 4.0856 – 0.0047 Age – 0.2223 Nchild + 0.0070 Nchild*Age
Male, unemployed:	Travel time = 3.8336 + 0.0001 Age – 0.2223 Nchild + 0.0070 Nchild*Age
Female, unemployed:	Travel time = 3.8935 – 0.0047 Age – 0.2223 Nchild + 0.0070 Nchild*Age

We can observe that the number of children has the same effect on the (log) travel time of employed and unemployed people and that there is no difference amongst genders. What the influence is of age and the number of children is exemplified in Table 10. In this table, we compared the average daily travel time for an individual of 30 years olds to that of a person of 40 years olds without children and to that of a 40 year old individual with 1 child and 2 children.

<INSERT TABLE 10 ABOUT HERE>

We can see that in general within the same age category males travel longer than females and employed people travel longer than their unemployed counterparts. Women travel less when growing older, while men travel somewhat more, and in general, an increasing number of children increases travel time accordingly, both for males and for females. This seems in contrast to the findings in Zwerts, *et al.* (2007), and therefore we tested for a significant effect of the interaction between the number of children and gender, however, this interaction variable clearly is not significant.

b. Travel Distance

After going through a similar procedure of first simplifying the fixed effects structure and then the random effects structure, the final model that is obtained looks as follows:

$$Y_{ij} = \beta_{0j} + \beta_{1j} age_{ij} + \varepsilon_{ij}$$

with $$\beta_{0j} = \varsigma_{00} + \varsigma_{01} nchild_j + \varsigma_{02} gender_{ij} + \varsigma_{03} HHinc_j + \varsigma_{04} employment_{ij} + u_{0j} \,,$$

$$\beta_{1j} = \varsigma_{10} + \varsigma_{11} nchild_j + \varsigma_{12} gender_{ij} + u_{1j} \,,$$

$$\varepsilon_{ij} \sim N(0, \sigma^2) \text{ and } \begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{10} & \sigma_1^2 \end{pmatrix}\right].$$

The AIC criterion equals 15085.5 for this final model, while -2*log-likelihood yields 15077.5. The significant fixed effect parameters can be found in Table 11.

<INSERT TABLE 11 ABOUT HERE>

Household income proves again to be a very significant variable. Giving the fact that all other variables remain constant, the lowest distance is traveled in households with the lowest income, while the one but highest incomes travel the longest distance.

Separating the equation for some categories might ease the interpretation, so therefore, for the highest income category, we split up according to gender and employment status:

Male, employed: Travel distance = 3.8273 – 0.0073 Age – 0.5732 Nchild + 0.0153 Nchild*Age

Female, employed: Travel distance = 3.8385 – 0.0153 Age – 0.5732 Nchild + 0.0153 Nchild*Age

Male, unemployed: Travel distance = 3.4481 – 0.0073 Age – 0.5732 Nchild + 0.0153 Nchild*Age

Female, unemployed: Travel distance = 3.4593 – 0.0153 Age – 0.5732 Nchild + 0.0153 Nchild*Age

It can be seen that the age and the number of children both have a negative effect on the number of kilometers driven, while their interaction variable has a positive effect. The slope for females on age is also steeper than that for males, indicating that the number of daily kilometers driven for females will decrease quicker with an increasing age than for males.

To be consistent with the discussed results on travel time, we will use the same example as in Table 10, but here we will write down the daily traveled distance for people in the highest income category, i.e. with a household income above 5000€ a month.

<INSERT TABLE 12 ABOUT HERE>

It can clearly be observed that on average females travel less kilometers than males. This is in line with the general knowledge that usually in a family the workplace for females is closer to home, so that they have to travel less. Employed people travel more than unemployed people, which sounds logical. The effect of age has been discussed earlier, and children seem to have a slightly increasing effect on the number of kilometers driven.

All these results are quite in accordance to what is generally expected.

5. Conclusions and Future Research

Most of the datasets collected in transportation research have 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 transportation literature is rather limited. Over the past twenty years, there has been an increasing interest in developing suitable techniques for the modeling and analysis of hierarchically structured data. In this paper, we exploit the most often and broadest class of models, i.e. the generalized mixed model strategy.

The results clearly show that there exists some correlation between individuals of the same household when modeling travel time and travel distance. Even up to 30% of the variation in travel time and/or distance 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 done too often and it needs to be dealt with in an appropriate way. The final mixed model that incorporates variables at both levels shows that the individual characteristics age, gender and employment status play an important role in modeling travel time and travel distance, where at household level also the number of children and (for travel distance) household income come into play.

In the presented models, all trips are considered at the same time. It will be interesting in the future to distinguish between different modes of travel and to do a separate analysis for each travel goal. It seems very logical that for work-related trips this correlation will be less, e.g. when compared to modeling leisure trips. Further research will take a closer look at these extensions of the presented models.

6. References

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), pp. 1231-1253.

Bryk, A.S. and Raudenbush, S.W. (1992) *Hierarchical Linear Models: Applications and Data Analysis Methods*, Newbury Park, CA: Sage.

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

Goulias, K.D. (2002) ‘Multilevel analysis of daily time use and time allocation to activity types accounting for complex covariance structures using correlated random effects’, *Transportation*, **29**(3), pp. 31-48.

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.

Kutner, M.H., Nachtsheim, C.J., Neter, J. and Li, M. (2004) *Applied Linear Statistical Models* (5th edition), McGraw Hill.

McNally, M.G. (2000) The activity-based approach, Center for Activity Systems Analysis, Paper UCI-ITS-AS-WP-00-4.

Molenberghs, G. and Verbeke, G. (2005) *Models for Discrete Longitudinal Data*, Springer-Verlag.

Ortúzar, J. de D. and Willumsen, L.G. (1994) *Modelling Transport*, Wiley, Chichester.

Schäfer, A. (2000) ‘Regularities in travel demand: an international perspective’, *Journal of Transportation and Statistics*, **3**(3), pp. 1-32.

Schäfer, A. and Victor, D. (2000) ‘The future mobility of the world population’, *Transportation Research A*, **34**(3), pp. 171-205.

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), pp. 323-355.

Verbeke, G. and Molenbergs, G. (2000) *Linear Mixed Models for Longitudinal Data*, Springer-Verlag.

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

Zwerts, E., Janssens, D. and Wets, G. (2007) ‘How the presence of children affects the parent's travel behavior.’ Paper accepted for presentation at the 86th Annual Meeting of the Transportation Research Board, January, 21-25, Washington D.C.

LIST OF TABLES

Table 1 Parameter Estimates (REML) of the Random Effects for the Unconditional Means Model on Travel Time

Parameter	**Estimate**	**Standard Error**
σ_H^2	0.2339	0.0163
σ^2	0.5398	0.0144

Table 2 Parameter Estimates (REML) of the Random Effects for the Unconditional Means Model on Travel Distance

Parameter	**Estimate**	**Standard Error**
σ_H^2	0.4735	0.0372
σ^2	1.2314	0.0342

Table 3 Parameter Estimates of the Fixed Effects for the Model on Travel Time at Level 2

Parameter	**Estimate**	**Standard Error**
Intercept	4.0794	0.1098
Nchild	-0.0301	0.0155
HHinc ≤ 750 €	-0.2941	0.1363
750 < HHinc ≤ 1875 €	-0.1931	0.1112
1875 < HHinc ≤ 3125 €	-0.1058	0.1107
3125 < HHinc ≤ 5000 €	-0.0244	0.1161

Table 4 Parameter Estimates (REML) of the Random Effects for the Model on Travel Time at Level 2

Parameter	**Estimate**	**Standard Error**
σ_H^2	0.2317	0.0162
σ^2	0.5393	0.0144

Table 5 Parameter Estimates of the Fixed Effects for the Model on Travel Distance at Level 2

Parameter	**Estimate**	**Standard Error**
Intercept	3.0533	0.2831
Nchild	-0.0901	0.0227
Ncars	0.2865	0.0380
HHinc ≤ 750 €	-0.6551	0.2168
750 < HHinc ≤ 1875 €	-0.2124	0.1670
1875 < HHinc ≤ 3125 €	-0.0098	0.1633
3125 < HHinc ≤ 5000 €	0.0480	0.1691
BTMH < 250 m	-0.3418	0.2201
250 ≤ BTMH ≤ 499 m	-0.2987	0.2204
500 ≤ BTMH ≤ 999 m	-0.2747	0.2208
1 ≤ BTMH ≤ 1.999 km	-0.1809	0.2244
2 ≤ BTMH ≤ 5 km	-0.0737	0.2334

Table 6 Parameter Estimates (REML) of the Random Effects for the Model on Travel Distance at Level 2

Parameter	**Estimate**	**Standard Error**
σ_H^2	0.4066	0.0346
σ^2	1.2253	0.0337

Table 7 Parameter Estimates of the Fixed Effects for the Model on Travel Time at Level 1

Parameter	Estimate	Standard Error
Intercept	3.5233	0.0533
Age	0.0043	0.0008
Gender = “male”	0.1136	0.0212
Employm. St. = “Employ.”	0.2166	0.0402

Table 8 Parameter Estimates of the Fixed Effects for the Model on Travel Distance at Level 1

Parameter	Estimate	Standard Error
Intercept	2.4140	0.0838
Age	0.0032	0.0013
Gender = “male”	0.2753	0.0322
Employm. St. = “Employ.”	0.4707	0.0648

Table 9 Parameter Estimates of the Fixed Effects for the Final Model on Travel Time

Parameter	**Estimate**	**Standard Error**
Intercept	3.8935	0.0748
Age	-0.0047	0.0013
Gender = ”male”	-0.0599	0.0510
Employm. St. = ”Employ.”	0.1921	0.0403
Nchild	-0.2223	0.0332
Nchild x Age	0.0070	0.0008
Gender = “male” x Age	0.0048	0.0012

Table 10 Average Daily Travel Times for Some Examples

Gender/employment	30 years, no children	40 years, no children	40 years, 1 child	40 years, 2 children
Male, employed	56.19 min.	56.24 min.	59.58 min.	63.12 min.
Female, employed	51.66 min.	49.28 min.	52.21 min.	55.31 min.
Male, unemployed	46.37 min.	46.41 min.	49.17 min.	52.09 min.
Female, unemploy.	42.63 min.	40.67 min.	43.09 min.	45.65 min.

Table 11 Parameter Estimates of the Fixed Effects for the Final Model on Travel Distance

Parameter	**Estimate**	**Standard Error**
Intercept	3.4593	0.1940
Age	-0.0153	0.0021
Gender = ”male”	-0.0112	0.0764
Employm. St. = ”Employ.”	0.3792	0.0607
Nchild	-0.5732	0.0511
HHinc ≤ 750 €	-0.8623	0.2106
750 < HHinc ≤ 1875 €	-0.3316	0.1599
1875 < HHinc ≤ 3125 €	-0.1185	0.1589
3125 < HHinc ≤ 5000 €	0.0227	0.1658
Nchild x Age	0.0153	0.0012
Gender = “male” x Age	0.0080	0.0018

Table 12 Average Daily Travel Distance for Some Examples

Gender/employment	30 years, no children	40 years, no children	40 years, 1 child	40 years, 2 children
Male, employed	36.90 km.	34.31 km.	35.66 km.	37.07 km.
Female, employed	29.36 km.	25.19 km.	26.19 km.	27.22 km.
Male, unemployed	25.26 km.	23.48 km.	24.41 km.	25.37 km.
Female, unemploy.	20.09 km.	17.24 km.	17.92 km.	18.63 km.