Transportation **29**: 31–48, 2002

Multilevel analysis of daily time use and time allocation to activity types accounting for complex covariance structures using correlated random effects

KONSTADINOS G. GOULIAS
Department of Civil and Environmental Engineering and The Pennsylvania Transportation Institute, The Pennsylvania State University, University Park, Pennsylvania, 16802, USA

Key words: activity-based, heterogeneity, longitudinal, multilevel, panel, travel

Abstract. In this paper multilevel analysis is used to study individual choices of time allocation to maintenance, subsistence, leisure, and travel time exploiting the nested data hierarchy of households, persons, and occasions of measurement. The multilevel models in this paper examine the joint and multivariate correlation structure of four dependent variables in a cross-sectional and longitudinal way. In this way, observed and unobserved heterogeneity are estimated using random effects at the household, person, and temporal levels. In addition, random coefficients associated with explanatory variables are also estimated and correlated with these random effects. Using the wide spectrum of options offered by multilevel models to account for individual and group heterogeneity, complex interdependencies among individuals within their households, within themselves over time, and within themselves but across different indicators of behavior, are analyzed. Findings in this analysis include large variance contribution by each level considered, clear evidence of non-linear dynamic behavior in time-allocation, different trajectories of change in time allocation for each of the four dependent variables used, and lack of symmetry in change over time characterized by different trajectories in the longitudinal evolution of each dependent variable. In addition, the multivariate correlation structure among the four dependent variables is different at each of the three levels of analysis.

1. Introduction

In recent years activity-based travel forecasting systems emerged as substitutes of and complements to more traditional travel demand forecasting systems (Pas 1999). For this, new conceptual and theoretical ideas have also been offered that provide frameworks for and models of time allocation to activities and travel. Activity-based forecasting systems require time allocation models that reflect or explicitly model activity participation such as work, shopping, visiting friends, traveling. Examples of these models can be the time expenditure for each activity type, the frequency of activity types in a day, the amount of time a person spends outside the home and at home, and so forth.

Four key modeling aspects that can ameliorate regression models for activity-based travel demand forecasting are considered in this paper. First, behavioral models can be improved if proper consideration of the contexts in which people allocate time is reflected in the models. For example, person-based models need to consider observed and unobserved household interactions. As argued and demonstrated by Golob and McNally (1997) the interaction in time use decisions within a household is of paramount importance in modeling travel behavior. Second, specification of these models can be improved when we incorporate observed and unobserved heterogeneity using models with more informative random structures (for transport examples see Bhat 2000 and Goulias 1999a). Third, for forecasting systems observed and unobserved longitudinal variation should be accounted for and explicitly represented in the models (for an example using latent class models see Goulias 1999b). Fourth, the usual regression model assumption of independent observations should be relaxed and the correlation among these observations analyzed and exploited in model formulation and interpretation.

Information about household interactions can be found in many existing databases that also contain other unexplored information about contextual behavior. For example, activity and travel diaries are collected from a few household members within a household and data on each trip, the person making the trip, and the household, are available. Data from diaries on persons within households may be viewed as data with a nesting hierarchical structure. In fact, single activity episodes and trips are nested within activity-trip chains (or tours), trip chains are nested within occasions of measurement, occasions are nested within individuals, individuals are nested within households, households are nested within neighborhoods, neighborhoods are nested within cities, cities are nested within regions, and so forth. Models that exploit this structure depict individual heterogeneity by explicitly considering the variation contributed by each context to behavior. In addition, quantitative estimates of the variation contributed by each context to a behavioral measure can be used to better understand individual behavior but also to improve model estimation.

One method to statistically account for the contextual behavior of individuals exploiting the hierarchical nature of these data is called multilevel analysis. The method uses strong correlations across observations at different levels in a hierarchy simultaneously without employing linear aggregation procedures that lead to biases. For example, analyses at the household level that aggregate behavioral indicators linearly, are substituted by person-based analyses that use the correlation of within household interactions of behavior. Multilevel analysis accounts for hierarchies in our data and may reflect, in a statistical sense, complex interaction among observations. In this way, testing for the existence of alternate correlation structures corresponding to alternate theories of travel behavior is possible and the relative variability contributed by each hierarchy level to the behavioral indicators can be quantified. Extending multilevel analysis to multiple indicators of behavior enables to also study compensatory movements in allocating time among different members in the same household and trading of time among different types of activities by a person. Moreover, this can be done with longitudinal data to study trajectories of change over time.

The models and the type of regression analysis used here is known by different names in different fields of research. For example, they have been named random coefficient models (Greene 1997: 669; Longford 1993), multilevel models (Goldstein 1995), mixed models (Searle et al. 1992), and hierarchical linear models (Bryk & Raudenbush 1992) among other labels attempting to describe the contextual nature of the data and/or the way of accounting for dependent variable variation from multiple sources. While some of the labels in this family of models may indicate subtle but important differences among the models, their key common characteristics are:

a) explicit recognition in model formulation of the hierarchical, multiple level and nested structure of the data to analyze, and
b) specification using three groups of regression components in the same regression model. The first group assumes constant sensitivity to explanatory variables among the units of analysis representing the mean effect of an explanatory variable on the dependent variable. The second group assumes a random deviation around this mean and the third group is the usual random error term(s) of the regression equation.

From among the different types of these models, a variant of repeated measures multilevel analysis focusing on time allocation to activities in a day is illustrated in this paper. Statistical models to study choices of individual time allocation to maintenance, subsistence, leisure, and travel time are estimated using nested hierarchies that include households, persons, and occasions of measurement. Parameters of heterogeneity, longitudinal and cross-sectional, at each hierarchical level, are estimated using random effects for each level. Random coefficients, that are associated with explanatory variables, are also estimated and their correlation with the context-specific random effects is studied. The paper contains two groups of model examples. In the first group of multilevel analysis, models that are very similar to the analysis of variance regression models are estimated. In the second group, we find models that depict trajectories of time allocation over the years of observation and account for the joint correlation structure of the four dependent variables considered here. In the next section the data used are described, followed by a section containing the models. The paper concludes with a summary.

2. Data used

The Puget Sound Transportation Panel (PSTP), which is the first general purpose urban household panel survey in the United States (Murakami & Waterson 1990), is used in this paper. In the panel, a survey is administered repeatedly on the same observations over time. Each survey, conducted at each point in time and called a wave, contains three groups of data that are household demographics, persons' social and economic information, and reported travel behavior in a two-day travel diary. Land use information around the place of residence and work have been added to each wave (details can be found in Goulias & Ma 1996) for a few waves. The data used in this paper are from the first five waves of the panel conducted in 1989, 1990, 1992, 1993, and 1994. The travel diaries include continuous 48-hour activities (excluding the in-home activities after the evening returning home) for each wave (for additional details see Murakami & Ulberg 1997). In the survey, each person was interviewed on the same two days in both waves and the diary includes every trip a person made during these two days. Each trip is described by trip purpose, mode, departure time, arrival time, travel duration, origin, destination, and distance traveled. Out-of-home activity engagement information can be derived using the trip purposes. The duration of an activity is computed by the difference between the start time of the next trip (departure from a given location) and the end time of the current trip (arrival at a given location) giving the sojourn time at an activity location.

In the original data set, trip purposes are classified into nine different types: work, school, college, shopping, personal business, appointments, visiting, free-time, and home during the day. Grouping of activities is needed in order to make the analysis tractable. Assuming that a person, within 24 hours, prioritizes his/her activity participation according to the relative importance of each activity, a natural grouping based on decreasing degree of constraint, importance, and spatial fixity is: subsistence (work, school, college), maintenance (shopping, personal business, appointments), leisure (visiting, free-time, home during the day) activities, and travel. For multilevel analysis, we retain the data of persons that participated in all five waves (stayers) and they have provided complete and consistent information in their questionnaires. They are 1201 persons in 758 households whose characteristics are provided in Table 1. The models in the next section consider the longitudinal measurements of total daily time expenditure in each activity type by each panel participant as nested within each individual and each individual as nested within her/his household (see Figure 1).

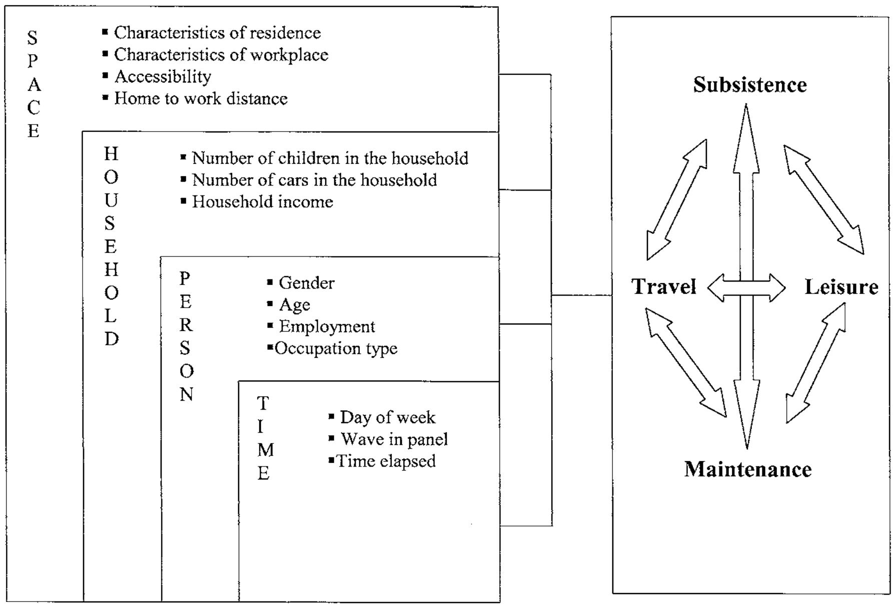

Figure 1. Pictorial representation of the multivariate and multilevel model.

3. Multilevel analysis

The following simple two level model illustrates a few key ideas underlying multilevel modeling. Consider a single dependent variable with two levels, a household and the person in it. A regression model with dependent variable Y with one explanatory variable X can be written as:

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

Y_{ij} is the dependent variable reported by a person i within a household j (with $i = 1, 2, \ldots,$ number of persons in household j, $j = 1, 2, \ldots,$ number of households in the sample). The variable X may be a person variable (e.g., age, gender, education, employment) or a household variable (e.g., number of vehicles owned, residential accessibility indicators). The random error term ε_{ij} is the usual regression error term defined at the lowest level, the person level, with $E(\varepsilon_{ij}) = 0$ and $var(\varepsilon_{ij}) = \sigma^2_{\varepsilon}$.

The coefficient β_{0ij} of this model has a mean γ_0 and the variation around this mean among households is depicted by the random variable u_{0j} with $E(u_{0j}) = 0$ and $var(u_{0j}) = \sigma^2_{u0}$ written as:

$$\beta_{0ij} = \gamma_0 + u_{0j}$$

The coefficient β_{1ij} associated with variable X can also be assumed as random:

$$\beta_{1ij} = \gamma_1 + u_{1j}$$

This coefficient has a mean γ_1 and its variation among households is depicted by a random variable u_{1j} with $E(u_{1j}) = 0$ and $\text{var}(u_{1j}) = \sigma^2_{u1}$. The two random variables u_{0j} and u_{1j} are also correlated with $\text{cov}(u_{0j}, u_{1j}) = \sigma^2_{u01}$. In this way, the effect of explanatory variable X on the behavior of an individual varies randomly from household to household and co-varies with the mean random effect among households. Along similar lines we can define variation at level 1 and add other covariance terms capturing interactions across levels. Hypothesis testing can then proceed to verify assumptions about interactions within a household and the effects of higher level effects on lower level behavioral indicators by testing the significance of variance and covariance terms. In the following section another lower level is added (occasions of measurement) and multiple regression equations are considered jointly. The estimates of the coefficients, variances, and covariances of the random error terms are used in behavioral interpretations to study heterogeneity within and among individuals and households.

Multiple dependent variable models (multivariate multilevel)

Figure 2 contains the basic equations of the model estimated here. In Equation 1, Y^T_{tij} is the amount of time dedicated to travel in a day of a wave, t, by a person i within a household j (with $t = 1, 2, 3, 4$, and 5, $i = 1, 2, \ldots$, number of persons in household j, $j = 1, 2, \ldots$, number of households in sample). The time points in this paper are the same for all individuals. Equations 2, 3, and 4 are defined in a similar way. These four equations are the level 1 model because they are written at the level of the time point (observation occasion). In all the examples here each time point is the first interview day in each panel wave.

The first term in the right hand side of each equation is a random intercept given by Equation 5. This component has specific meaning. For example, β^q_{tij} is the daily time expenditure of person i in household j at time t for activity q when all other explanatory variables are zero (called the initial status herein). The term ε^q_{tij} is a random temporal variation (also called within person variation) and it is the deviation of time expenditure around γ^q_0. The term u^q_{ij} is a random person to person variation (also called within household variation) and it is also a deviation of time expenditure around γ^q_0. The term v^q_j is a random household to household variation and it is also a deviation of time expendi-

$$Y^T_{tij} = \beta^T_{tij} + \beta^T_1 \text{ Time}_{tij} + \beta^T_2 \text{ Male}_{ij} + \gamma^T_1 \text{ Employed}_{tij} + \gamma^T_2 \text{ Log}(HW)_{tij} + \gamma^T_3 \text{ Adults}_{ij} + \gamma^T_4 \text{ Age}_{tij} + \gamma^T_5 \text{ Time}^2_{tij} \quad (1)$$

$$Y^S_{tij} = \beta^S_{tij} + \beta^S_1 \text{ Log}(HW)_{tij} + \gamma^S_2 \text{ Employed}_{tij} + \gamma^S_2 \text{ Male}_{ij} + \gamma^S_3 \text{ Age}_{tij} + \gamma^S_4 \text{ Age65 } p_{tij} + \gamma^S_5 \text{ Friday}_{tij} \quad (2)$$

$$Y^M_{tij} = \beta^M_{tij} + \gamma^M_1 \text{ Time}_{tij} + \gamma^M_2 \text{ Time}^2_{tij} + \gamma^M_3 \text{ Employed}_{tij} + \gamma^M_4 \text{ Male}_{ij} + \gamma^M_5 \text{ Age65 } p_{tij} + \gamma^M_6 \text{ Log}(HW)_{tij} + \gamma^M_7 \text{ K6} - 17_{tj} \quad (3)$$

$$Y^L_{tij} = \beta^L_{tij} + \beta^L_1 \text{ Time}_{tij} + \gamma^L_1 \text{ Employed}_{tij} + \gamma^L_2 \text{ Driver}_{ij} + \gamma^L_3 \text{ K6} - 17_{tij} + \gamma_4{}^L \text{ K6} - 17 \text{ Male}_{tij} + \gamma^L_5 \text{ Log}(HW)_{tij} + \gamma^L_6 \, PD_{tij} + \gamma^L_7 \text{ 2 Cars}_{tij} + \gamma^L_8 \text{ 3 PCars}_{tij} + \gamma^L_9 \text{ Monday}_{tij} \quad (4)$$

$$\beta^q_{tij} = \gamma^q_0 + v^q_j + u^q_{ij} + \varepsilon^q_{tij} \quad \text{where} \quad q = T, S, M, L \quad (5)$$

$$\beta^T_1 = \gamma^T_1 + v^T_{1j}, \quad \beta^T_2 = \gamma^T_2 + u^T_{2ij}, \quad \beta^S_1 = \gamma^S_1 + v^S_{1j}, \quad \beta^L_1 = \gamma^L_1 + v^L_{1j}, \quad (6)$$

Figure 2. The multivariate multi-equation model.

ture around γ^q_0. These are also called random error components and they are assumed normally distributed with $E(\varepsilon) = E(u) = E(v) = 0$, with $\text{Var}(\varepsilon) = \sigma^2_\varepsilon$, $\text{Var}(u) = \sigma^2_u$ and $\text{Var}(v) = \sigma^2_v$ to be estimated. In Equations 1, 2, 3, and 4 all the gamma coefficients are defined in a similar way as in a typical regression model. The βs, however, are defined as random with a mean and a variation around their means γs. This variation can be due to the temporal, personal, and/or household levels. In this way, we can define a variety of equations at each of these levels to represent heterogeneous behavior that is either due to temporal fluctuations, personal variation, or household variation. For example, in Equation 1 the second and third coefficients are given by Equation 6 together with the second coefficient of Equation 3 and the second coefficient of Equation 4. Two of them are random varying coefficients associated with time (rate of change in time allocation for travel and for leisure) from household to household, which in essence means time allocation varies across households for travel time and leisure in a non-linear way with calendar time. One coefficient associated with gender for travel varies randomly among persons, which means the effect of gender on travel time allocation is different from person to person. The last random coefficient is associated with the logarithm of the distance between home and work in the subsistence equations. This means different households will be sensitive in different ways, when allocating time to subsistence, to their distance between home and work. For example, households with adults in one type of profession may have a large positive effect (they live far from work and they spend long hours at work) and households with adults in another type of profession may have a negative effect (they live far from work and they spend fewer hours at work).

At each level we have a level-specific variance-covariance matrix of all the random terms (εs, *u*s, *v*s). The significance of the elements in each of the three matrices can be tested using goodness-of-fit measures based on the deviance, which is the difference in the –2 Log(likelihood) at convergence between two nested (in terms of specification) models. In addition, the γs can also be tested if they are significantly different than zero using a *t*-test. The γs in Equations 1 to 6 are called the *fixed effects* and the remaining terms are called the *random effects* at each of the three levels in the hierarchy.

Estimation of all the fixed and random parameters can be accomplished either by Full Information Maximum Likelihood, FIML, applied to *Y* directly or applied to the least-squares residuals, called Restricted Maximum Likelihood-REML that can be used in tandem with a generalized least squares approach. Longford (1993), Bryk and Raudensbush (1992), and Goldstein (1995) provide a comprehensive review of estimation techniques, their performance assessment, and detailed algorithms. In this paper Godstein's (1995)

Table 1. Average sample characteristics in this study (standard deviation in parentheses).

Variable	1989	1990	1992	1993	1994
Travel time (minutes/day) by a person	87.3 (60.8)	84.2 (56.6)	79.4 (55.1)	81.3 (59.3)	82.2 (59.8)
Subsistence (minutes/day) by a person	301.6 (254.8)	311.1 (253.2)	295.5 (257.7)	292.5 (258.1)	286.3 (265.7)
Maintenance (minutes/day) by a person	60.6 (97.7)	46.0 (79.4)	45.8 (73.3)	41.9 (68.6)	48.37 (83.8)
Leisure (minutes/day) by a person	120.0 (159.2)	105.8 (155.9)	103.7 (155.5)	109.7 (158.0)	99.5 (157.8)
Age	46.7 (13.3)	47.9 (13.7)	50.0 (13.7)	51.0 (13.7)	52.1 (13.7)
No. of children ages 1 to 5 in household	0.213 (0.53)	0.200 (0.51)	0.158 (0.48)	0.147 (0.46)	0.133 (0.48)
No. of children ages 6 to 17 in household	0.437 (0.80)	0.440 (0.80)	0.450 (0.82)	0.438 (0.80)	0.433 (0.80)
Numbers of cars in the household	2.34 (1.1)	2.34 (1.1)	2.36 (1.1)	2.27 (0.97)	2.26 (0.98)
Percent employed in household (%)	69.8	69.4	73.9	65.8	64.5

Iterative Generalized Least Squares (IGLS) approach is used that separates estimation of the fixed from the random parameters at different steps in sequence repeatedly until no change is observed in the estimates in subsequent steps. Goldstein (1995), has also improved the IGLS algorithm when based on FIML using a modified Iterative GLS called RIGLS. In fact, this method provides standard errors of coefficient estimates that are conservative (larger) and for this leading to more parsimonious models. In an experiment performed in this study the IGLS and RIGLS gave similar results and identical conclusions about significance of variables. In addition, experiments with non-linear duration models also produced similar results.

The first model presented here contains no explanatory variables (Table 2). It is called the null model or fully unconditional model and it is used as a benchmark to assess other model specifications that include explanatory variables and regression coefficients (fixed and/or random) at each level. The estimation method used is Goldstein's RIGLS. The first portion of Table 2 shows the lion's share in the proportion of variance is within persons and across time points for travel, maintenance, and leisure. This is followed by the proportion of variance across persons (except for subsistence in which the person level has the highest proportion because it is driven by employment status). Households contribute the smallest proportion of variation for travel, subsistence, and leisure. The key finding here is that the household level variance, is more than 1/3 of the person level variance and therefore not negligible. The multilevel specification appears to be justified. Similar findings characterize all four dependent variables considered indicating we should specify models using explanatory variables capturing and/or depicting individual and household change and person and household characteristics that may not change over time. In a separate analysis, not shown here, it was also found that a similar to the household magnitude of variation is contributed by the place of residence. However, including both the household and the household residence level as two separate levels was not possible and considerable confounding of effects exists between household and place of residence (with the indicator of the place of residence given by the census tract within which the household resides). For this reason, explanatory variables capturing the effects of the place of residence need to also be included. As we will see later this level-specific variance can be decomposed within each level to other sources. Increasing the number of explanatory variables within each level, however, affects the variance of other levels when there is considerable interaction across levels. It is rare to eliminate the unexplained variation within each level completely because of unobserved significant factors and measurement error. The bottom half of Table 2 contains the estimated covariances (below the diagonal) and the estimated correlation coefficients within each of the three levels for the combination of the four dependent variables.

Table 2. Multivariate error components models.

Model component	Travel time		Subsistence time		Maintenance time		Leisure time	
Fixed effect	Coef.		Coef.		Coef.		Coef.	
Grand mean (γ_0)	82.8	(1.1)	297.3	(6.6)	48.5	(1.4)	106.8	(3.1)
Random effects	(σ^2)	Percent	(σ^2)	Percent	(σ^2)	Percent	(σ^2)	Percent
Temporal variation within persons (ε_{0ij})	2,549.3	74.8	26,485.3	39.7	5,731.5	86.5	18,353.2	74.1
Person variation within households (u_{0ij})	606.1	17.8	30,640.3	45.9	434.4	6.6	4,127.4	16.7
Between households variation (v_{00j})	254.9	7.4	9,588.0	14.4	461.9	6.9	2,273.5	9.2
Total	3,410.3	100.0	66,713.6	100.0	6,627.8	100.0	24,754.1	100.0
−2 Log L				290,898.8				

Variance/covariance matrices (upper triangle correlations)

	Between occasions				Between persons				Between households			
	Travel	Subs.	Maint.	Leis.	Travel	Subs.	Maint.	Leis.	Travel	Subs.	Maint.	Leis.
Travel	2,549.3	−0.027	0.221	0.225	606.1	0.421	−0.293	−0.158	254.9	−0.152	0.432	0.278
Subs.	−218.5	26,485.3	−0.269	−0.244	1,815.7	30,640.3	−0.907	−0.652	−238.1	9,588.0	−0.433	−0.190
Maint.	846.4	−3313.9	5,731.5	−0.004	−150.1	−3,308.2	434.4	0.606	148.2	−910.3	461.9	0.236
Leis.	1,539.1	−5373.5	−44.4	18,353.2	−249.5	−7,334.9	811.3	4,127.4	211.4	−886.1	241.6	2,273.5

The lowest values of correlation coefficients are found in the "Between Occasions" quadrant in which the four highest correlation coefficients are also much larger than their standard errors. Strong negative correlation is shown between subsistence and the other two activity types (maintenance and leisure). This pattern is repeated at the person and household levels displaying a clear pattern in which participation in more work and/or school activities leads to the inhibition of participation in other activities within each person, within each household, and across households when explanatory variables are absent. Across persons there is a strong negative correlation between maintenance and subsistence, which is also negative and strong at the household level indicating the separation of tasks among household members that "specialize" in work-school versus "shopping." The differences in the correlation structures across the three levels indicate complex and dissimilar trade-offs taking place within each of the three contexts. The correlations at the person level are mostly driven by individual activity participation and the correlations at the household level are mostly driven by the people within each household and their "agreement" in time allocation. The next model offers some additional insights.

When we include explanatory variables in this repeated measures model we have a variety of options in a wide spectrum of models. On one end of the spectrum, we can have explanatory variables defined at level 1 alone that are multiplied by regression coefficients that do not vary along the time, person, or household dimensions. On the other end of the spectrum, we can have explanatory variables that are defined at all levels and their coefficients can vary within any of the levels. These coefficients can also be functions of other variables to reflect the systematic variation in regression coefficients with respect to one or more temporal, person, or household characteristics. This very wide array of options allows us to specify models that are quite flexible and are very interesting for studying the dynamics in time allocation of individuals and account for individual and group heterogeneity as well as their interdependence.

The results of Table 2, extensive analysis using single equation models (models that consider serial correlation and non-normal random error terms) and a variety of multivariate model specifications are combined in the model presented in Tables 3 and 4. This is a multivariate model that contains additional explanatory variables, a covariance structure reflecting the relationship among the four dependent variables (travel, subsistence, maintenance, and leisure) and variances at each of the three levels (time, person, and household) that are themselves functions of explanatory variables. In this way the heteroskedasticity of random components' within each of the three levels and the cross-correlations of these random components within each level have been accounted for in the model.

The travel model has an estimated mean initial status of 77.29 minutes of

Table 3. Multivariate multilevel model (standard errors in parentheses).

Fixed effects	Travel	Subsistence	Maintenance	Leisure
Initial status	77.29 (5.52)	194.11 (16.67)	81.33 (3.05)	98.10 (12.46)
Time (in months)	–0.31 (0.11)		–0.90 (0.17)	–0.24 (0.08)
Time squared × 10	0.04 (0.02)		0.12 (0.03)	
Interview day is Monday (=1, 0 othewise)				–14.75 (5.79)
Interview day is Friday (=1, 0 otherwise)		–135.78 (62.46)		
Employed (=1, 0 otherwise)	–7.55 (2.58)	197.91 (8.68)	–16.04 (3.53)	–28.89 (6.90)
Male (=1, 0 otherwise)	6.50 (1.82)	54.15 (6.70)	–13.43 (2.18)	
Age	–0.18 (0.08)	–2.23 (0.31)		
Age 65+ (= 1, 0 otherwise)		–47.46 (10.72)	12.49 (3.63)	
Have a driver's license (= 1, 0 otherwise)				35.03 (11.29)
Number of children 6–17 years old			2.92 (1.42)	16.51 (3.73)
Number of children 6–17 years old (w/male)				–16.13 (4.67)
Number of adults in household	2.26 (0.41)			
Log (work-home distance in miles)	11.47 (0.96)	43.18 (3.32)	–5.69 (1.27)	–13.43 (2.58)
Residence population density (persons/mile^2)				2.69 (0.78)
Indicator for 2 cars (= 1, 0 otherwise)				12.92 (6.12)
Indicator for 3+ cars (= 1, 0 otherwise)				20.10 (6.67)

travel for unemployed and approximately 8 minutes less for employed persons. In addition, the logarithm of work-home distance has a large positive effect on travel as expected. The average monthly change rate is a decrease of 0.31 minutes, which in the 60 months of the panel yields a significant decrease

Table 4. Multivariate multilevel model (random effects and variance/covariance matrix elements).

Model component[a]	Travel time	Subsistence time	Maintenance time	Leisure time
Random effects[b]	(σ^2)	(σ^2)	(σ^2)	(σ^2)
Time	0.040			0.632
Log (work-home distance)		392.29		
Gender	469.5			

Variance/covariance matrices[a] (upper triangle correlations)

	Between occasions				Between persons				Between households			
	Travel	Subs.	Maint.	Leis.	Travel	Subs.	Maint.	Leis.	Travel	Subs.	Maint.	Leis.
Travel	2,515.8	–0.052	0.232	0.238	123.6				295.4		0.481	0.301
Subs.	–401.1	23,952.4	–0.269	–0.247		9,454.8	–0.881	–0.560		565.6		
Maint.	877.2	–3,138.9	5,688.9			–1,015.2	140.3	0.542	161.5		381.2	
Leis.	1,596.5	–5,112.7		17,843.3		–2,979.4	350.9	2,990.4	203.2			1,540.8
–2 Log *L*						288,769.7						
	Deviance from error components model (Table 2) = 2,129.1											

Notes

[a] All components of variance/covariance reported are larger than twice their standard error.

[b] Deviance from a model with all four variances set to zero = 73.3.

of approximately 19 minutes. We also observe a non-linear and significant effect that is captured by the coefficient of the time squared. The combined effect of these two variables produces a minimum for the travel time of 71 minutes reached at approximately 40 months into the study and then a gradual increase over time. In addition, males travel on average approximately 6.5 minutes more than females and for each year of a person's aging we observe a decrease of 0.18 minutes of travel, which is considerably less than the effect of one additional adult in the household. The number of children in the household, car ownership, day of the week when interview was completed, and drivers' license do not influence significantly travel time in this model formulation (similar findings are obtained in other specifications using single equation models).

In addition, in the travel model the rate of change is different across different households and the variance of the random component associated with this is 0.04. Moreover, the effect of gender is different among different individuals in the sample with variance equal to 469.5. The estimates of these variances have standard errors that are at least twice smaller than their estimates. As shown in the second part of Table 4 when compared to the second part of Table 2 all variances are much smaller due to the addition of explanatory variables in the specification, which is as desired because we have eliminated some of the unexplained variation.

The subsistence model has an estimated mean initial daily allocation of 194 minutes for unemployed and approximately 392 minutes for employed persons, with males allocating 54 minutes more than females. The logarithm of work-home distance has a large positive effect in a similar way as for travel indicating persons living far from work also tend to spend longer hours at work. In addition, on Fridays on average people tend to allocate 136 less minutes to subsistence than on other days of the week (e.g., leaving work early for the weekend). The average speed of change is a decrease of 2.23 minutes shown by the age coefficient, which in the 60 months of the panel yields more than 10 minutes. In addition, the age group of 65 and older has on average 45 less minutes of daily allocation to subsistence as expected due to retirement. In addition, the effect of the home to work distance is different across households possibly reflecting the differential effect among households.

The maintenance model has estimated initial daily allocation of 81 minutes for unemployed persons and approximately 16 minutes less for employed persons, with males allocating 13 minutes less than females. The logarithm of work-home distance has a modest negative effect while the presence of children in the age group 6 to 17 years old affects positively the amount of time dedicated to maintenance activities with each child contributing 3 additional minutes per day. People over 65 years old dedicate on average 12 additional minutes per day than the other age groups. The average speed of change is a substantial decrease of 0.90 minutes, which in the 60 months of the panel would decrease the daily allocation by 54 minutes. However, the model also contains a second order time effect term that when combined with the speed term shows maintenance time reaching its minimum at the 38th month and then going up again.

The leisure model has estimated mean initial daily allocation of 98 minutes for unemployed persons and approximately 29 minutes less for employed persons, with no difference between males and females. In addition, on a Monday people are more likely to dedicate on average 15 minutes less to leisure presumably due to the weekend. Persons with a driver license allocate 35 minutes more than non-drivers to leisure activities. The average monthly change rate is a decrease of 0.24 minutes. In the 60 months of the panel it implies a modest decrease of 14 minutes that is approximately 0.5 times smaller than the difference between employed and unemployed people but still not negligible. In a household each child in the age group 6 to 17 years old contributes approximately 17 minutes to each survey participant's daily leisure activities.

The effect of children within individuals has an average value of 16.5 minutes per child age 6 to 17 in the household. In this model, the mean effect of children on leisure time daily expenditure is different between men and women (the variable Gender is coded 0 for women and 1 for men) showing that in essence children add time to the average leisure daily duration of women alone. This effect was also found not to vary randomly among households or persons. The population density around the residence and car ownership are also significant positive contributors to the amount of time allocated to leisure activities.

All the random effects components are not zero statistically, which indicates we have residual variance that can be explained by explanatory variables not included in the model, unexplainable factors at each level, and measurement error. In addition, the three random error terms are also heteroskedastic. The correlation structure among the dependent variables is similar to the error components model of Table 2 with fewer significantly different than zero covariances. Also, from the variance components we see significant variation exists among households for travel and leisure changing rates from household to household. When considering the multivariate model as a whole, the total variance at the household level is 1,286, the variance at the person level is a 3,232, and the variance within the persons' is 12,500.

4. Summary and conclusions

In this paper, three key factors (behavioral context, heterogeneity, and longitudinal variation) with high potential in improving regression equations for activity-based travel demand forecasting systems are considered. The analysis focuses on time allocation to activities in a day using statistical models to study choices of individual time allocation to maintenance, subsistence, leisure, and travel time. The implicit hierarchy in a panel survey with travel diaries has been exploited to quantify the effect of context at the levels of households, persons, and occasions of measurement. Behavioral heterogeneity is also quantified using components of variation for each context and random coefficients, associated with explanatory variables, within each context. Longitudinal variation is examined using single dependent variable and multiple dependent variable formulations each providing distinct information about the role of calendar time and age in time allocation. The analysis presented here provides unique insights about time allocation and demonstrates how multilevel analysis may prove to be a powerful tool for travel behavior analysis. This is, however, the first time that multivariate multilevel models are used in transport analysis. A few earlier applications are the introduction of multilevel and contextual philosophy in Ma and Goulias (1997), an error components four-level binary analysis for mode choice constraints in Goulias (1999b), and the discrete choice models accounting for spatial clustering in Bhat (2000).

The models estimated show for travel time, maintenance, and leisure the amount of variance within persons and across time points are larger than the variance across persons. However, for subsistence the person level is the highest due to employment status. The key finding here, however, is that the household level variance, is more than 1/3 of the person level variance and therefore not negligible. In addition, the relationship among the unobserved components of these four dependent variables is different in each of the three levels considered in this paper.

The models in this paper provide clear evidence of non-linear dynamic behavior in time-allocation when examined longitudinally. Longitudinal trajectories that are followed by the four dependent variables are dissimilar with two of them (travel and maintenance) following U shaped functions of time and the other two (subsistence and leisure) linear decay functions of time. All dependent variables are driven by explanatory variables that are also non-linear functions of time. The correlations at each of the three levels considered here are dissimilar indicating lack of symmetry in the longitudinal evolution of each dependent variable.

The analysis here offers further support in favor of the feasibility and appropriateness in using multi-equation approaches to travel behavior analysis. There are, however, some limitations worthy of mention. Throughout this study we did not account for the intervening effect of the number of episodes for each activity type (e.g., number of distinct trips made in a day, number of times a person enters and stays at a store, etc.). Using the same dataset, however, a within a day study was done in Ma and Goulias (1998). In addition, explanatory variables of change such as relocation, purchase of a new car, birth of another child, death of a family member were not used. This was taken into account partially in the study here by including a few time varying variables. However, indicators of explicit change are not included and their effects cannot be shown clearly. There are other limitations related to the estimation of the random components and their reliability that are discussed extensively elsewhere (Goldstein 1995; Bryk & Raudenbush 1992). Finally, the nature of the dependent variables may be considered to be a limited dependent or similar to episode durations. This would require different estimation procedures (see Golob & McNally 1997). Preliminary trials using single equation estimates with error terms that are gamma distributed did not lead to different conclusions about the significance of coefficient estimates. In addition, very little is known about violation of assumptions in multilevel models and research on many aspects of the models presented here is very active.

Acknowledgements

Portion of this paper was prepared while the author was on sabbatical leave under the Fulbright/FLAD Chair in Knowledge Management and Policies at the Instituto Superior Téchnico da Universidade Técnica de Lisboa, Portugal. Partial funding was also provided by the Center for Intelligent Transportation Systems (CITranS) at the College of Engineering, The Pennsylvania State University. Three anonymous referees offered valuable insights and they are gratefully acknowledged.

References

Bhat C (2000) A multi-level cross-classified model for discrete response variables. *Transportation Research* 34B: 567–582.

Bryk AS & Raudenbush SW (1992) *Hierarchical Linear Models*. Newberry Park: Sage.

Goldstein H (1995) *Multilevel Statistical Models.* London and New York: Edward Arnold.

Goulias KG (1999a) Multilevel random effects analysis of modal use constraints and perceptions on public transportation using data from Germany. In: Sacharov LJ (ed) *Urban Transport V, Urban Transport and the Environment for the 21st Century* (pp 181–190). Southampton: WIT Press.

Goulias KG (1999b) Longitudinal analysis of activity and travel pattern dynamics using Generalized Mixed Markov Latent Class Models. *Transportation Research* 33B(8): 535–557.

Goulias KG & Ma J (1996) Analysis of longitudinal data from the Puget Sound transportation panel. Task B: Integration of PSTP databases and PSTP codebook. Final Report. # 9619. The Pennsylvania Transportation Institute, The Pennsylvania State University, University Park, PA.

Golob TF & McNally MG (1997) A model of activity participation and travel interactions between household heads. *Transportation Research* 31B: 177–194.

Greene WH (1997) *Econometric Analysis*, Third Edition. New Jersey: Prentice Hall.

Longford NT (1993) *Random Coefficient Models.* Oxford: Clarendon Press.

Ma J & Goulias KG (1997) An analysis of activity and travel patterns in the Puget Sound Transportation Panel. In: Ettema DF & Timmermans HJP (eds) *Activity-Based Approaches to Travel Analysis* (pp 189–207) The Netherlands: Pergamon.

Ma J & Goulias KG (1998) Forecasting Home Departure Time, Daily Time Budget, Activity Duration, and Travel Time Using Panel Data. Preprint 98-00392 TRB CD-ROM. Paper presented at the 77th Annual Transportation Research Board Meeting, January 11–15, Washington DC.

Murakami E & Watterson WT (1990) Developing a household travel panel survey for the Puget Sound Region. *Transportation Research Record* 1285. 40–48.

Murakami E & Ulberg C (1997) The Puget Sound transportation panel. In: Golob TF, Kitamura R & Long L. *Panels for Transportation Planning Methods and Applications* (pp 159–192) Boston: Kluwer.

Pas EI (1999) Time in travel choice modeling: from relative obscurity to center stage. In: Garling T, Laitila T & Westin K (eds) *Theoretical Foundations of Travel Choice Modeling.* Pergamon, Amsterdam.

Searle SR, Casella G & McCulloch CE (1992) *Variance Components*. New York: Wiley.

About the author

Konstadinos Goulias is associate professor of Civil Engineering and program director at the Pennsylvania Transportation Institute at Penn State University. He is the chair of the A1C04 committee on Travel Behavior and Values of the Transportation Research Board and is the recipient of the 1999 Fulbright/FLAD Chair.