A decade of longitudinal travel behavior observation in the Puget Sound region: sample composition, summary statistics, and a selection of first order findings

K. G. Goulias*,
Professor, The Pennsylvania State University

N. Kilgren,
Senior Planner, the Puget Sound Regional Council

and
T. Kim
Research Assistant and Ph.D. Candidate, The Pennsylvania State University

*Corresponding author.
201 Transportation Research Building, Mid-Atlantic Universities Transportation Center, The Pennsylvania Transportation Institute, The Pennsylvania State University, University Park, PA, 16802. Tel: 1-814-863-7926. Fax: 1-814-865-3039, Email: goulias@psu.edu.

Paper Submitted to the 10th International Conference on Travel Behaviour Research,
Moving through nets: The physical and social dimensions of travel
Lucerne, 10-14 August 2003

Version - June 8, 2003

Introduction

Activity-based approaches to travel demand forecasting are rapidly moving into practice. These approaches have increased the number and complexity of the typical travel behavior indicators (e.g., number of trips made by purpose, distance traveled, mode used), adding many time allocation variables such as departure from home, amount of time traveling, and amount of time dedicated to activities to the task of data collection (Axhausen, 1998) and to data analysis (Pendyala, 2003). They also amplified our need to identify relatively homogenous behavioral patterns representing observed behavior. On one hand, the identification of a small set of behavioral patterns is a first step needed in creating algorithms for the prediction of behavior (Ma, 1997, Arentze and Timmermans, 2000). On the other hand, forecasting can be accomplished best using longitudinal information about changes in the lives of persons and their households and changes in their activity and travel behavior patterns (Goulias and Kitamura, 1997). These two aspects, i.e., concise summaries of behavior and longitudinal dependency, are combined in this paper using a unique database that allows us to identify activity and travel patterns and their temporal evolution for an unprecedented long time span, from 1989 to 2000, in a large American metropolitan area, using a substantially large sample (a little over 1700 households per time point of observation). To do this we employ a pattern recognition technique known as cluster analysis.

Developing a finite set of activity and travel pattern types requires some kind of pattern recognition technique that summarizes the values of observed variables and groups them into clusters. The most popular procedure uses persons from a diary survey as the elementary units (observations). For each person we have the values (scores) for variables such as number of trips made in a day, amount of time spent traveling to work, amount of time spent in activity participation, and so forth. Then, a statistical or other data analytic software is used to identify groups of observations that have reported values for these variables close to each other and classify the observations into clusters of similar behavior. Closeness is decided based on some sort of difference among criteria variables (e.g., the values of these variables or the values of derivative variables). Cluster membership may be crisp and deterministic (each person belongs to a single cluster with probability equal to one) or it can be fuzzy and probabilistic (each person belongs to a cluster with a probability of membership that may be less than one). The usual pattern recognition technique in activity analysis and travel behavior has been the k-means statistical technique, which is crisp, deterministic, and derived for continuous variables.

In travel behavior analysis, pattern recognition for cluster development has been employed since the early applications of activity-based approaches. These clusters were used to summarize human behavior in terms of several relatively homogeneous groups using statistical-mathematical algorithms designed for continuous variables (Pas, 1982, 1983, see also Ma, 1997 for a review of others). A subsequent step along this research path was to also study changes in activity and travel patterns within a day or a week (Pas and Koppelman, 1987, Pas, 1988) by observing changes in the selection of clusters. Later, building on these efforts, longer term dynamic aspects of travel behavior, using data from the Puget Sound Transportation Panel (PSTP), were studied, using k-means, by Ma and Goulias, 1996, 1997a, 1997b. This was accomplished by separating activity from travel and by examining the switching among clusters from one day to the next and from one year to the next in the panel. In that study, the idea of clusters emerging as the outcome of multilevel processes at the household, person, and other external to the decision making unit levels was also introduced and analyzed in more detail. Using the same clusters, Goulias (1999) also developed a model system in which transitions among clusters over time were studied employing latent class analysis methods and demonstrated the existence of multiple paths of change (that he called longitudinal heterogeneity). Variation among these same clusters from multiple levels was also examined using regression models designed to incorporate the multilevel data structure explicitly (Goulias and Kim, 2001). In a parallel study, Krizek (2003), using a subset of the PSTP data, developed "lifestyle" clusters using the K-means statistical routine. All these studies used a limited time frame and clustering methods that are designed for continuous variables (i.e., the k-means routine). The cluster membership information is then used as a categorical dependent variable and a variety of explanatory variables are employed as determinants of cluster classification. In the past few years, however, substantial progress has been made in clustering and classification algorithms, aided by inexpensive fast computing. In this paper we analyze all nine waves available in PSTP using one of these newer algorithms with the following objectives:

a) derive an optimal number of clusters that summarize activity and travel behavior between 1989 and 2000 in the PSTP data;
b) describe these clusters in terms of daily activity participation and travel behavior;
c) identify social and demographic groups most likely to belong to each cluster; and
d) identify any time effects on these clusters.

Two sets of models are presented in this paper. The first set examines the issues of traveling alone or with others and the second examines the time allocated to activities and travel. The remainder of the paper begins with an overview of the panel and the data used here. This is followed by a description of the clustering method and the data analysis results. The paper concludes with a summary.

The Puget Sound Transportation Panel

The Puget Sound Transportation Panel (PSTP) was designed as a "general purpose" urban household panel survey tailored to transportation analysis. The PSTP was also created as a tracking device of changes in employment, work characteristics, household composition, vehicle availability, travel behavior and responses to changes in the transportation environment, and attitudes and values (Murakami and Watterson, 1990). As Murakami and Ulberg (1997) state, the objectives of PSTP are:

"to monitor changes in household composition, location and employment characteristics; to monitor changes in travel behavior and responses to changes in the transportation environment; and to examine the effects of changes in attitudes and values on mode choice and travel behavior." (Murakami and Ulberg, 1997, page 163)

The PSTP data used here are a record of travel behavior aiming to represent approximately 3.3 million residents (based on data from the US Census of 2000) in Seattle and its surroundings. The survey started in 1989 and continues until today in the four counties (King, Kitsap, Pierce, and Snohomish) of the Puget Sound region in the Northwest corner of the continental US. In each wave a household questionnaire and a two-day travel diary are administered on essentially the same households and their members 15 years or older. In this way, we accumulate households that participated at multiple time points. Unlike more traditional transportation cross-sectional surveys, PSTP takes similar measurements (i.e., surveys) repeatedly on the same observations over time. Each wave of the PSTP includes a travel survey that collects information on household demographics, person social and economic circumstances, and reported travel behavior on two consecutive days for each person in the recruited household that is 15 years or older to capture driving age individual behavior. The PSTP currently available data are from nine travel surveys in the years 1989, 1990, 1992, 1993, 1994, 1996, 1997, 1999, and 2000. This provides a database with unique capability for regional forecasting models.

Forecasting models that use panel surveys as their source of data have the potential to reproduce observations much better than their cross-sectional counterparts because they use cross-sectional variation and they are grounded on observed individual and household longitudinal histories allowing us to study individual and household changes. Moreover, panels allow us to test if the cross-sectional demand models are valid representations of behavioral trends in the region. For example, one can estimate a series of models from each year of the panel and test if the relationship between travel behavior and demographic variables remains constant over time. In addition, other models can be estimated to examine if behavior is changing in a linear or non-linear way with time and if trip making is most likely to increase, decrease, or even go through cycles while controlling for all other exogenous factors.

Table 1 provides a summary of the evolution for a few key household composition characteristics in the waves 1 to 9 panel participants. Table 2 displays a fairly constant sample composition in terms of number of males in each household, employed persons outside the home, number of drivers in the household, and car ownership. Table 3 contains some key travel behavior indicators that are of particular importance for typical travel demand forecasting models.

Since 1997 the original objectives of PSTP were also enriched by the introduction of questions on traveler information system use and telecommunications and computer ownership and use. For example, wave 7 (1997) *"was specifically designed to provide insight into how people use data (traveler information) to make transportation decisions"* (Kilgren, 1998). To accomplish this, PSRC, with funding from the Washington State Department of Transportation and the U.S. DOT, incorporated a few supplemental questions about Intelligent Transportation Systems (ITS – the combination of telecommunications and information technology to aid transportation system management and operations). A tenth wave was started in late Fall 2002 and data from this survey are expected in mid-2003 or soon thereafter. The data used in this paper are from the first 9 waves.

The tables above give us a general overview of the types of information contained in the PSTP and the change in the average values of key indicators. A somewhat more informative overview can be obtained when relatively homogeneous groups of behavior can be created and then correlated to other factors. Before moving to the analysis some background information about the clustering technique employed is provided.

Latent Class Cluster Analysis

The technique selected to identify groups of homogeneous patterns of activity and travel behavior in the first decade of the PSTP database is *latent class cluster analysis*. This technique:

a) includes a J-category latent variable, each category representing a cluster
b) uses many "dependent" or clustering variables (named *criteria* variables herein);
c) uses a mixture of multiple types of criteria variables (e.g., continuous, categorical, ordered, count);
d) uses and tests the effect of covariates of many different types;
e) is more flexible than many other clustering algorithms;

f) is a model-based clustering approach, so it provides probabilistic membership of observations in clusters; and
g) provides convenient interpretable output

In this paper we use notation and model formulation similar to Vermunt and Magidson (2002). Assume there is one latent variable (X) representing the time allocation of a person during the two observation days in the PSTP waves. Different categories of this variable X denote different types of activity-travel behavior and the probability of belonging to each category of variable X represents the proportion of persons that choose that specific type of time allocation. Using observed data we would like to identify how many distinct groups we have, find the proportion of persons in each group, and gain insights about their temporal evolution.

For each person in our sample we observe M measures (indicators) of activity and travel behavior indicated by the symbol Y that can be used to infer membership in the categories of the latent variable **X**. A third set of variables, which are not included as criteria variables in the clusters, are used as explanatory variables and for this indicated with the symbol **Z**.

The probability density of the Ys given a set of **Z** values is:

$$f(Y \mid Z) = \sum_{x} \pi(X \mid Z) f(Y \mid X, Z) \tag{1}$$

where $\pi(X \mid Z)$ is the probability of belonging to a certain latent class given a set of covariate values.

If the Y variables belonging to different clusters (categories of variable X) are assumed to be mutually independent given the latent class and the covariates, we obtain:

$$f(Y \mid Z) = \sum_{x} \pi(X \mid Z) \prod_{m=1}^{M} f(Y_m \mid X, Z) \tag{2}$$

Since the scores on the latent variable given the covariates are assumed to come from a multinomial distribution, probability of belonging to a given latent class can be calculated as follows:

$$\pi(X \mid Z) = \frac{e^{\eta_{X|Z}}}{\sum_{X} e^{\eta_{X|Z}}} \tag{3}$$

where the term η is a linear combination of the main effects of the latent variable (γ_{x_j}) and the covariate effects on the latent variable $(\gamma_{z_l x_j})$ defined as:

$$\eta_{X|Z} = \sum_{j=1}^{J} \gamma_{x_j} + \sum_{l=1}^{L} \sum_{j=1}^{J} \gamma_{z_l x_j} \tag{4}$$

One way to visualize this model is to consider a cross-classification table underlying the model in which latent and observed variables are included. This table has dimensions equal to the categories of all the variables when all variables are categorical. The cell values of this table are the entities we are trying to estimate using formulations as in Equation 4. As in many latent class models the likelihood function takes the familiar form shown below where θ denotes the unobserved parameters to be estimated.

$$LogL = \sum_{i} n_i \log f(Y_i \mid Z_i, \theta) \tag{5}$$

The parameters in equation 5 can be estimated by the Expectation Maximization (EM) algorithm, which produces Maximum Likelihood estimates under specific conditions. In the examples here we use the Vermunt and Magidson (2002) method, which is a combination of EM with Newton-Raphson. Standard errors for the parameter estimates are computed using the Hessian matrix (matrix of the second order derivatives of the estimating equation). As the number of parameters to estimate increases, the degrees of freedom decrease rapidly and for this reason we run into a variety of operational problems such as identification (inability to compute a parameter) or lack of convergence (subsequent estimation step parameters are not close enough). Most latent class models are also sensitive to local maxima of the likelihood function used in estimation, which can be circumvented by testing multiple models using different initial trial values for the parameters (see also Goulias, 1999). Estimation of models of this type is in essence a hierarchical iterative process in which:

a) We start with a one cluster assumption and a simple model is estimated;
b) experimentation proceeds by increasing the number of clusters until identification is no longer possible for some parameters, the cluster sizes become too small to be meaningful, and the difference in goodness of fit between successive models is not significant; and
c) selecting one or more models that appear to be a reasonable description of the observed data, we define alternate modeling options such as correlations among criteria variables and variances within each cluster starting another iterative cycle. This goes on until the addition of a more complex structure no longer yields a significant improvement (for nested models we can use a formal statistical step as a stop criterion).

Within these three steps we also have two additional "mini-steps." For each model we first develop starting values for the unknown parameters we are estimating that are drawn from a distribution of randomly selected moments. For a given set of starting values we perform maximum likelihood iterations first using the EM algorithm until the values of subsequent iterations reach a predefined difference (or the total number of EM iterations reaches a maximum number). Then, the algorithm switches to a Newton-Raphson algorithm until again a predetermined convergence criterion value is reached or the maximum number of iterations is reached. In this way, we can exploit advantages of both algorithms, i.e. the stability of EM even when far away from the optimum and the speed of Newton-Raphson when close to the optimum (Vermunt and Magidson, 2002).

In the same way as other latent class models, statistical goodness-of-fit measures for latent class cluster models are the typical chi-square statistics used also in the cross-categorical data analysis. The first measure is the likelihood ratio chi-square, G^2 or $L^{2.}$ It has a chi-square distribution with degrees of freedom given by the number of "free" parameters (Total number of different response patterns - the number of estimated model parameters – 1). It represents the opposite of an R^2 in regression because it is the amount of unexplained variation by the model. Therefore, higher values indicate models that do not fit the data well and lower values represent better fitting models. When models are nested, (i.e., they differ only in the number of estimated parameters), we could create the difference between the G^2 of the two models. This difference is chi-square distributed and can be used for hypotheses testing. This cannot be done between models that differ in the number of clusters because they are not nested. Based on L^2, the Bayes information criterion (BIC), Akaike information criterion (AIC) and the Consistent Akaike information criterion (CAIC) are computed to measure goodness of fit and to take into account model parsimony penalizing models with many parameters. The lower the BIC, AIC or CAIC values, the better the model we estimate (McCutcheon, 2002).

The approach followed in the analysis presented in this paper has some advantages over the more popular cluster analysis using the k-means and then using some kind of regression to identify the composition of each cluster. First, the latent class cluster method for identifying clusters is designed for combinations of continuous and discrete criteria variables while the k-means method is defined for continuous variables only. Second, the method used here allows for a probabilistic membership of each observation in each cluster. This provides flexibility in observation classification while the k-means does not allow for that. Third, post-processing of the cluster data using regression is not required because the method used allows the inclusion of covariates. There are other advantages of latent class methods in general and the specific implementation used here that are discussed in Vermunt and Magidson, 2002.

Data Analysis

The basic ingredients of an activity based approach for travel demand analysis (Jones, Koppelman, and Orfeuil, 1990 and Arentze and Timmermans, 2000) are: a) explicit treatment of travel as derived demand; b) the household is the fundamental decision making unit; and c) explicit consideration of constraints. Manheim (per Arentze and Timmermans, 2000), discussed the idea of travel demand as derived demand first in the context of a transportation system and its simulation. In that context, participation in activities (e.g., work, shop, leisure) motivates travel but travel could also be an activity as well (e.g., taking a drive). These activities are viewed as episodes (starting time, duration, and ending time) and they are arranged in a sequence forming a pattern of behavior that can be distinguished from other patterns (a sequence of activities in a chain of episodes). In addition, these events are not independent and their interdependency is accounted for in the theoretical framework. The second aspect is also related to this last point. Considering the household is considered to be the fundamental social unit (decision making unit), the interactions among household members can be explicitly modeled to capture task allocation and roles within the household. The episodes mentioned above are then tasks within a larger "project" in which each person in the household is involved. This brings up another issue that is the relationships among household members and the concomitant task allocation. In addition, relationships change as households move along their life cycles and the individual's commitments (e.g., escorting a child to school) and constraints (e.g., need to be at a specific time at a specific place) change and these are depicted in the activity-based model. These constraints in their temporal, spatial, and social dimensions are also receiving increasing attention in activity-based models. For example, using the time-space prism idea from Hagerstrand (1970), Pendyala, (2003), uses regression models to model the size of the action space within the prism and Arentze and Timmermans (2000) insert reflections of these constraints in the form of model parameters and/or rules in a production system format.

The inputs to these models are the typical regional model data of social, economic, demographic information of potential travelers to create schedules followed by people in their everyday life. Land use information could also be used when available at the detail required to develop individual behavioral models. The output are detailed lists of activities pursued, times spent in each activity, and travel information from activity to activity (including travel time, mode used, and so forth). This output is very much like a “day-timer” for each person in a given region.

Preceding the creation of an activity-based system for forecasting is another step. In this step we need to understand the formation of patterns in allocating time to activities and travel and the evolution of these patterns over longer periods (e.g., periods that span many years). This is needed because many forecasting applications are designed for a **twenty** year horizon. Panel data of the type described in this paper help us to develop this understanding as illustrated below with two examples: a) traveling alone or with others; and b) allocation of time to activities and travel.

Traveling alone and with others

In activity-based approaches research is particularly needed on group decision-making and interactions among persons. For example, to depict the exchange of resources in terms of a network of relationships among people can be depicted by using indicators of their interaction. In this way one can build *social networks* that provide the structural environment within which opportunities and constraints to individual action (e.g., activity participation and travel) are shown explicitly. While this conceptualization has not been used in data collection and modeling, the interaction among household members has been recognized as an important factor affecting behavior (Van Wissen, 1989, Golob and McNally, 1996, Chandrasekharan and Goulias, 1999, Gliebe and Koppelman 2002) and we build on these initial efforts in the data analysis of this section considering the following variables:

1. Number of trips traveling alone with car/truck/sport utility vehicle (SUV) in day 1 and day 2
2. Number of trips traveling with relatives with car/truck/SUV in day 1 and day 2
3. Number of trips traveling with others non-relatives with car/truck/SUV in day 1 and day 2
4. Number of trips with car/truck/SUV for which others were involved, but their relationship with the respondent is unknown in day 1 and day 2.

To derive clusters of behavior we start from a one cluster model and build in sequence models with more clusters until estimation is no longer possible due to lack of identification for one or more parameters. The explanatory variables used here are at the person, household, and temporal levels. Differences among individuals in traveling together are determined by personal and household characteristics and for this reason we use employment and age as explanatory variables. One of the most important variables for household task allocation is the number of children and to reflect this we use the total number of children 1 to 5 years old in the household at the time of the interview as well as the number of children in the age group 6 to 17. A variable that is of paramount importance in traveling is car ownership and this is the third household-level variable used here. The county of residence and the sample type are used to control for the initial selection of the households when recruited to participate. In addition, the year of the wave is a covariate used to account for and study changes over time. The year when the person started in the panel is also used to account for the different composition of the replenishment of observation units sample (refreshment samples) and possible panel fatigue.

Table 5 provides a summary of this sequence until the solution with 5 clusters is reached. The six-cluster solution produced estimation problems with variables that could not be identified. Subsequent refinements and additions of explanatory variables to the five-cluster solution improved the goodness-of-fit statistics substantially as the last two rows of the table show. The model of the final step is the model we will examine in more detail here. Table 6 describes the five clusters in terms of the persons' behavior within each cluster. The first cluster of behavior and the largest (31.6% of sample) is the cluster of persons that make most of their trips alone with car/truck/SUV in both days. The second group of almost equal size (30.3% of sample) appears to make more trips overall with car/truck/SUV, and similar to the first cluster an almost equal amount of trips alone. This group is also characterized by a substantial number of trips with relatives (2.1 in day 1 and 2.2 in day 2). The three remaining clusters have much smaller size of membership. The third cluster is characterized by a large number of trips with relatives (other household members) and a small number of other trip types. In contrast, the fourth cluster is characterized by a large number of trips solo and a relatively large number of trips with others who are not relatives. The last cluster has very few trips per day.

Cluster composition can also be examined in terms of the social and demographic characteristics of the sample unit. This is accomplished in two ways: a) examining the significance of each covariate in determining the probability of belonging to each cluster (in essence testing the significance of the γs in equation 4); and b) given that the covariates are significantly affecting a cluster, examining the cluster composition in terms of the covariate's values. Table 7, in essence containing estimated conditional probabilities of membership, reports both aspects. Employment is a significant factor for all clusters. On one hand, employed persons most often select clusters 1, 2, and 4, in that order of popularity. On the other hand, unemployed persons are spread throughout the five clusters with the highest percentages in cluster 3, followed by clusters 2, 1, 5, and 4. Interestingly, the very young participants to the survey tend to travel most often with others, selecting 3 and 4 as the most popular clusters for them. These are the clusters with the highest percentages for traveling with relatives (cluster 3; note also that persons in this cluster tend not to travel alone) and for traveling with others (cluster 4, which, however, also features more solo travel). A substantial amount of this group (22%) appears to travel little using car/truck/SUV. The next two age groups appear to be more likely to follow travel behavior as in cluster 1 (solo travelers) as the preferred one. As expected the age group 35-44 travel~~s~~ the most and select~~s~~ cluster 2 that contains many trips with relatives (presumably young children as the next portion of the table also shows). The age groups 45-54 and 55-64 show a clear majority in cluster 1 with second in cluster 2, which again is most likely motivated by the presence of children in the household (e.g., the probabilities for children 6 to 17 for cluster 2 that are between 40% and 52%). Older individuals populate most clusters with a dissimilar tendency to the very young (except for clusters 3 and 5).

Figure 1 provides a depiction of changes from 1989 to 2000 using the coefficients γs in equation 4. Although the figure exaggerates the overall increase in cluster 1 it is also indicative of a general trend away from the other clusters and toward cluster 1. Average estimated conditional probabilities are reported immediately underneath the figure and show a clear increase in cluster 1 with a variety of interesting fluctuations over the years of the survey, which are not equally spaced in time. Similar analysis was also performed using the year at which a person entered the panel as one of the explanatory variables. Very few indicators for this variable were found significantly different than zero, indicating that only for one or two years in each wave we may have a significantly different choice of clusters.

Time allocation to activities and travel

In the second cluster analysis using the PSTP data we demonstrate another set of group identification using the following variables:

1. total number of trips in each of the two days (Tottrip1, Tottrip2),
2. total daily travel time to work (Wttime1, Wttime2),
3. total daily travel time to non-work places (Nwttime1, Nwttime2),
4. total amount of time spent in activities per day (Atime1, Atime2).

Table 8 provides an overview of the model estimation results as estimation moved toward the final five cluster model that is selected for illustration here. Other five cluster models with additional explanatory variables did not yield a converged solution that satisfied all the convergence criteria. Table 9 shows that we have five types of activity-travel behavior. Table 10 shows the average membership probability for employment, age, number of children ages 1 to 5 and number of cars in the household.

The first cluster is characterized by the absence of any employed persons and it is populated by the very young and the seniors in our sample (see also Figure 2). A few of them may work and this is reflected in the small amount of travel to work. This group is also characterized by the relatively lower amount of time spent in activities outside the home base in both days of the survey (4-5 hours per day). The other four clusters are dominated by the employed persons and there is no clear and definite trend with respect to the age of the panel participants.

Figure 3 shows the values of the γs for the time effect on cluster membership (Equation 4). At first glance, this may appear to be a sinusoidal pattern. This pattern, however, is not translated into a sinusoidal pattern of the probability of cluster membership as a function of time because of other factors affecting this probability and the values of the γs of the other probabilities. In fact, Figure 4 shows a fairly stable evolution in favor of cluster 1 (the most popular).

Summary

Data from a decade collected in the Seattle region have been analyzed in this paper using one of the latest methods in pattern recognition called latent class cluster analysis. Both data and method are insightful about travel behavior and appear to provide unique opportunities for travel behavior research. Using all nine waves of the PSTP that are currently available two types of analysis were performed and illustrated here.

The first analysis examines patterns of travel interactions among persons by developing clusters based on the number of trips traveling alone with car/truck/sport utility vehicle (SUV) in each of the two days of the survey, number of trips traveling with relatives with car/truck/SUV in each of the two days of the survey, and number of trips traveling with others that are non-relatives or for whom we have no information. The five cluster solution appeared to be the best using these variables. These are made by two larger clusters containing approximately 30% of the respondents each and three smaller clusters containing 10% to 14% of the respondents each. Analysis of the role and significance of covariates in cluster classification showed that as expected employment is a key shaper of travel behavior. In addition, different age groups opt for a different pattern of traveling alone and with others in a day and the number of children in households play a key and important role. All this confirms past research on travel behavior and serves as a simple verification. Most important, however, is the finding of a general temporal shift to travel patterns of more traveling alone trips and less traveling with others.

In the second analysis example time allocation is examined by first building clusters using as criteria variables the total number of trips in each of the two days (Tottrip1, Tottrip2), the total daily travel time to work (Wttime1, Wttime2), the total daily travel time to non-work places (Nwttime1, Nwttime2), and total amount of time spent in activities per day (Atime1, Atime2). Unlike the first analysis in this case we have one large cluster (more than 30% of the observations) populated by unemployed persons. Then we have a second cluster containg approximately 23% of the sample and two clusters containing approximately 19% of the sample each. The last cluster contains only 7% of the observations. Three clusters (2, 3, and 4) are the clusters of the employed persons and the differ both in terms of activity and travel allocations as well as person and household characteristics. In terms of time evolution we observe a strong stability in the cluster membership with a small increase in the membership of the first cluster.

Both analyses did not yield major longitudinal shifts when we consider membership for each cluster at each time point. The frequencies within each time point, however, are the sum of many possible opposing shifts (also called marginal frequencies). As seen elsewhere when analyzing panel data this is the combined result of possibly a considerable amount of changes at the person level and it may mask a variety of longitudinal behaviors as illustrated in Goulias, 1999. That type of analysis is left as a future task.

Acknowledgments

Funding for this paper was provided by the Puget Sound Regional Council in Seattle and the Mid-Atlantic Universities Transportation Center at the Pennsylvania State University.

References

Arentze T. and H. Timmermans (2000) Albatross A Learning Based Transportation Oriented Simulation System. European Institute of Retailing and Services Studies, Technical University of Eindhoven, Eindhoven, NL.

Axhausen K.W. (1998) Can we ever obtain the data we would like to have? In *Theoretical Foundations of Travel Choice Modeling* (eds Garling T., T. Laitila, K. Westin), Pergamon. Pp.305-323.

Chandrasekharan B. and K.G. Goulias (1999) Exploratory longitudinal analysis of solo and joint trip making in the Puget Sound transportation panel. *Transportation Research Record*, 1676, pp. 77-85.

Gliebe J. P. and F. S. Koppelman (2002) A model of joint activity participation between household members. *Transportation* 29(1). Pp 49-72.

Golob T. F. and M.G. McNally (1996) A model of household interactions in activity participation and the derived demand for travel. Paper presented at the 75th Annual Transportation Research Board Meeting, Washington, D.C.

Goulias K.G. (1999) Longitudinal analysis of activity and travel pattern dynamics using generalized mixed Markov latent class models. *Transportation Research*, Part B, 33(8), pp. 535-557.

Goulias K. G. and T. Kim (2001) Multilevel analysis of activity and travel patterns accounting for person- and household-specific observed and unobserved effects simultaneously. *Transportation Research Record*, 1752, pp. 23-31.

Goulias, K. G., and R. Kitamura (1997) Regional travel demand forecasting with dynamic microsimulation models In *Panels for Transportation Planning: Methods and Applications* (eds Golob, Kitamura, Long). Kluwer. Chapter 13, pp. 321-348.

McCutcheon, A.L. (2002) Basic concepts and procedures in single- and multiple-group latent class analysis. In Applied Latent Class Analysis (Edited by J. A. Hagenaars and A. L. McCutcheon), Cambridge University press, Cambridge, UK..

Jones P., F. Koppelman, and J Orfeuil (1990) Activity analysis: State-of-the-art and future directions. In *Developments in Dynamic and Activity-Based Approaches to Travel Analysis. A compendium of papers from the 1989 Oxford Conference* (ed. P. Jones). Avebury, UK. Pp. 34-55.

Kilgren N. (1998) Transportation Panel Survey (Wave7): Overview of the Puget Sound Transportation Panel Survey (1989-1997). In http://www.nawgits.com/psrc/index.htm (accessed November, 2002).

Krizek, K.J. (2003) Planning, household travel, and household lifestyles. In *Transportation Systems Planning: Methods and Applications*, ed. K.G. Goulias, CRC Press LLC. Boca Raton, FL, 2003.

Ma J. (1997) An Activity-Based and Micro-Simulated Travel Forecasting System: A Pragmatic Synthetic Scheduling Approach. Unpublished Ph.D. Thesis. Department of Civil and Environmental Engineering. The Pennsylvania State University, University Park, PA.

Ma, J. and K. G. Goulias (1996) Multivariate marginal frequency analysis of activity and travel patterns in the first four waves of the Puget Sound transportation panel. *Transportation Research Record*, 1556, pp. 67-76.

Ma, J. and K. G. Goulias (1997a) A dynamic analysis of activity and travel patterns using data from the Puget Sound transportation panel. *Transportation*, 24(1) , pp.1-23.

Ma J. And K. G. Goulias (1997b) An analysis of activity and travel patterns in the Puget Sound transportation panel. Chapter 10 In *Activity-Based Approaches to Travel Analysis* (Editors D.F. Ettema and H. J.P. Timmermans). Pergamon Elsevier Science LTD, The Netherlands. pp. 189-207.

Murakami E. and Wattterson W.T. (1990) Developing a household travel panel survey for the Puget Sound region. *Transportation Research Record* 1285, pp.40-46.

Murakami E. and Ulberg C. (1997) The Puget transportation panel. In *Panels for Transportation Planning: Methods and Applications* (eds. T.F. Golob, R. Kitamura, and L. Long). Kluwer, Boston, MA. Pages 159-192.

Recker W. W. and M. G. McNally (1985) Travel/Activity Analysis: Pattern Recognition, Classification and Interpretation. *Transportation Research A* Vol. 19A, No.4, 1985, pp. 279-296.

Recker W.W., M. G. McNally, and G. S. Root (1986) A Model of Complex Travel Behavior: Part I - Theoretical Development. *Transportation Research A* Vol. 20A, No. 4, pp.307-318.

Pas E.I. (1982) Analytically Derived Classifications of Daily Travel-Activity Behavior: Description, Evaluation, and Interpretation. *Transportation Research Record*, 897, pp. 9-15.

Pas E.I. (1983) A Flexible and Integrated Methodology for Analytical Classification of Daily Travel-Activity Behavior. *Transportation Science*, Vol. 17, No. 4, pp. 405-429.

Pas E. I. and F. S. Koppelman (1987) An Examination of the Determinants of Day-to-Day Variability in Individuals' Urban Travel Behavior *Transportation* Vol. 14, no. 1, pp. 3-20.

Pas E. I. (1988) Weekly Travel-Activity Behavior. *Transportation*. Vol. 15, nos. 1-2, pp. 89-109.

Pendyala R. (2003) Time use and travel behavior in space and time. Chapter 2 in *Transportation Systems Planning: Methods and Applications*. Edited by K.G. Goulias, CRC Press, Boca Raton, FL, pp. 2-1 to 2-37.

Van Wissen L. (1989) A model of household interactions in activity patterns. Paper presented at the International Conference on Dynamic Travel behavior Analysis. July 18-19, Kyoto University Hall, Kyoto, Japan.

Vermunt J. K. and Magidson J. (2002) Latent class cluster analysis. In Applied Latent Class Analysis (edited by J. A. Hagenaars and A.L. McCutcheon), Cambridge University Press, Cambridge, UK. pp.89-106.

Table 1 Evolution of household composition in PSTP

Year of wave		Number of children younger than 6	Number of children 6 to 17 years old	Number of adults (18 and older)	Household size
1989	Mean	0.23	0.44	1.92	2.59
	N	1712	1712	1712	1712
	Std. Deviation	0.57	0.83	0.65	1.26
1990	Mean	0.22	0.45	1.91	2.57
	N	1793	1793	1793	1793
	Std. Deviation	0.53	0.82	0.61	1.25
1992	Mean	0.19	0.44	1.90	2.53
	N	1569	1569	1569	1569
	Std. Deviation	0.52	0.84	0.61	1.25
1993	Mean	0.19	0.44	1.89	2.52
	N	1900	1900	1900	1900
	Std. Deviation	0.52	0.83	0.63	1.26
1994	Mean	0.18	0.44	1.89	2.51
	N	1750	1750	1750	1750
	Std. Deviation	0.51	0.83	0.61	1.25
1996	Mean	0.16	0.42	1.93	2.52
	N	1730	1730	1730	1730
	Std. Deviation	0.48	0.83	0.66	1.25
1997	Mean	0.16	0.41	1.90	2.47
	N	1982	1982	1982	1982
	Std. Deviation	0.48	0.82	0.66	1.27
1999	Mean	0.14	0.42	1.95	2.51
	N	1774	1774	1774	1774
	Std. Deviation	0.46	0.83	0.73	1.30
2000	Mean	0.12	0.42	1.86	2.40
	N	1701	1701	1701	1701
	Std. Deviation	0.42	0.85	0.66	1.27
Total	Mean	0.18	0.43	1.90	2.51
	N	15911	15911	15911	15911
	Std. Deviation	0.50	0.83	0.65	1.26

N = number of households

Table 2 Time evolution of key household characteristics

Year of wave		Number of persons employed outside the home	Number of men in household	Number of drivers in household	Number of household vehicles
1989	Mean	1.34	0.93	1.85	2.12
	N	1712	1712	1712	1712
	Std. Deviation	0.86	0.56	0.74	1.13
1990	Mean	1.26	0.92	1.83	2.15
	N	1787	1787	1787	1786
	Std. Deviation	0.82	0.55	0.70	1.10
1992	Mean	1.30	0.90	1.81	2.23
	N	1593	1593	1593	1579
	Std. Deviation	0.80	0.55	0.69	1.21
1993	Mean	1.20	0.91	1.82	2.06
	N	1906	1906	1906	1906
	Std. Deviation	0.82	0.58	0.71	1.08
1994	Mean	1.17	0.92	1.82	2.10
	N	1771	1771	1771	1771
	Std. Deviation	0.82	0.57	0.72	1.19
1996	Mean	1.20	0.92	1.80	2.09
	N	1755	1755	1755	1753
	Std. Deviation	0.83	0.57	0.68	1.08
1997	Mean	1.16	0.94	1.79	2.09
	N	1992	1992	1992	1990
	Std. Deviation	0.86	0.58	0.75	1.14
1999	Mean	1.24	0.92	1.81	2.13
	N	1795	1795	1795	1792
	Std. Deviation	0.95	0.60	0.78	1.15
2000	Mean	1.23	0.89	1.78	2.10
	N	1707	1707	1707	1705
	Std. Deviation	0.93	0.60	0.76	1.13
Total	Mean	1.23	0.92	1.81	2.12
	N	16018	16018	16018	15994
	Std. Deviation	0.86	0.58	0.73	1.14

N = number of households

Table 3 Time evolution of key personal travel behavior characteristics

Year of wave		Number of home based trips to work in day 1	Number of home based trips to college in day 1	Number of home based trips to school in day 1	Number of home based trips to shopping in day 1	Number of home based other trips in day 1	Number of other-to-other trips in day 1	Number of work-to-other trips in day 1	Number of returning home trips in day 1	Total number of trips in day 1
1989	Mean	0.48	0.04	0.03	0.20	0.72	1.38	0.32	1.46	4.64
	N	3390	3390	3390	3390	3390	3390	3390	3390	3390
	Std. Deviation	0.53	0.21	0.19	0.44	0.86	1.76	0.59	0.85	2.83
1990	Mean	0.51	0.03	0.04	0.17	0.67	1.40	0.29	1.40	4.51
	N	3497	3497	3497	3497	3497	3497	3497	3497	3497
	Std. Deviation	0.56	0.20	0.19	0.41	0.85	1.89	0.56	0.86	2.94
1992	Mean	0.49	0.03	0.04	0.16	0.71	1.19	0.25	1.41	4.26
	N	3059	3059	3059	3059	3059	3059	3059	3059	3059
	Std. Deviation	0.55	0.16	0.20	0.39	0.87	1.64	0.51	0.85	2.72
1993	Mean	0.49	0.05	0.04	0.15	0.66	1.18	0.25	1.37	4.19
	N	3684	3684	3684	3684	3684	3684	3684	3684	3684
	Std. Deviation	0.55	0.23	0.20	0.38	0.85	1.69	0.52	0.85	2.78
1994	Mean	0.47	0.03	0.04	0.14	0.67	1.22	0.27	1.34	4.19
	N	3461	3461	3461	3461	3461	3461	3461	3461	3461
	Std. Deviation	0.55	0.18	0.20	0.38	0.84	1.76	0.53	0.85	2.85
1996	Mean	0.47	0.02	0.04	0.16	0.72	1.24	0.25	1.38	4.27
	N	3425	3425	3425	3425	3425	3425	3425	3425	3425
	Std. Deviation	0.55	0.13	0.19	0.39	0.88	1.76	0.52	0.86	2.83
1997	Mean	0.50	0.03	0.04	0.13	0.67	1.23	0.24	1.36	4.21
	N	3939	3939	3939	3939	3939	3939	3939	3939	3939
	Std. Deviation	0.56	0.17	0.20	0.36	0.87	1.74	0.50	0.87	2.83
1999	Mean	0.47	0.01	0.04	0.15	0.72	1.34	0.22	1.36	4.30
	N	3535	3535	3535	3535	3535	3535	3535	3535	3535
	Std. Deviation	0.55	0.12	0.19	0.38	0.89	1.86	0.48	0.89	2.91
2000	Mean	0.45	0.02	0.04	0.10	0.67	1.26	0.22	1.26	4.02
	N	3259	3259	3259	3259	3259	3259	3259	3259	3259
	Std. Deviation	0.55	0.15	0.20	0.32	0.87	1.83	0.49	0.82	2.85
Total	Mean	0.48	0.03	0.04	0.15	0.69	1.27	0.26	1.37	4.29
	N	31249	31249	31249	31249	31249	31249	31249	31249	31249
	Std. Deviation	0.55	0.17	0.19	0.38	0.87	1.78	0.52	0.86	2.84

Table 4 Overview of the first set of dependent variables in the PSTP

Year of wave		Total number of trips in day 1	Total Number of Trips in day 2	Total travel time spent on work or work related in day 1	Total travel time spent on work or work related in day 2	Total travel time spent on non-work in day 1	Total travel time spent on non-work in day 2	Total activity time in day 1	Total activity time in day 2
1989	Mean	4.64	4.45	17.83	17.03	65.4789	65.9171	463.59	455.08
	N	3390	3390	3390	3390	3389	3390	3373	3378
	Std. Deviation	2.831	2.834	24.837	21.878	53.00078	55.72875	246.569	259.988
1990	Mean	4.51	4.44	22.64	22.13	60.1068	60.0215	437.90	444.57
	N	3497	3494	3497	3494	3494	3492	3460	3458
	Std. Deviation	2.941	2.916	35.632	34.772	50.08691	48.45403	251.607	259.569
1992	Mean	4.26	4.17	19.96	19.72	58.4881	60.8330	435.33	428.69
	N	3059	3059	3059	3059	3059	3059	3030	3033
	Std. Deviation	2.718	2.781	30.335	29.865	47.12299	54.97418	253.778	261.326
1993	Mean	4.19	4.22	20.96	19.66	57.7212	60.2172	428.38	434.40
	N	3684	3684	3684	3684	3684	3683	3656	3649
	Std. Deviation	2.783	2.857	33.911	30.677	50.06286	55.15724	256.872	264.544
1994	Mean	4.19	4.09	21.42	19.30	58.4331	59.4837	423.22	407.92
	N	3461	3461	3461	3461	3461	3461	3433	3433
	Std. Deviation	2.854	2.882	35.880	34.856	49.95830	52.67749	256.626	263.905
1996	Mean	4.27	4.19	19.66	19.16	61.9828	60.8669	429.43	425.47
	N	3425	3425	3425	3425	3424	3425	3401	3384
	Std. Deviation	2.827	2.893	31.293	30.166	54.55503	53.94596	257.543	266.860
1997	Mean	4.21	3.97	21.39	20.09	60.1165	58.0970	430.17	417.57
	N	3939	3939	3939	3939	3939	3938	3901	3895
	Std. Deviation	2.829	2.724	33.086	32.164	53.46647	50.01524	257.860	264.582
1999	Mean	4.30	4.12	19.88	18.99	61.3950	62.0331	421.21	415.32
	N	3535	3535	3535	3535	3532	3534	3493	3490
	Std. Deviation	2.911	2.899	29.202	28.974	53.92807	57.45074	256.185	260.670
2000	Mean	4.02	3.86	19.95	19.54	58.2782	57.3639	401.92	398.40
	N	3259	3259	3259	3259	3257	3256	3242	3224
	Std. Deviation	2.847	2.819	32.292	31.810	51.82419	51.99771	254.468	262.659
Total	Mean	4.29	4.17	20.45	19.53	60.2262	60.5049	430.15	425.30
	N	31249	31246	31249	31246	31239	31238	30989	30944
	Std. Deviation	2.845	2.850	32.081	30.854	51.72007	53.44034	255.142	263.232

Table 5 The sequence of models estimated for traveling alone and with others

Model Type	L^2 (or LL)	BIC	Number of parameters
1-Cluster	-307685	615451.8	8
2-Cluster	-278165	556732.6	39
3-Cluster	-264053	528831	70
4-Cluster	-253793	508631.8	101
5-Cluster	-247495	496355.6	132
6-Cluster*	-244159	490005.3	163
5-Cluster**	-243490	488511.8	148
5-Cluster***	-242526	486830.7	172

* produced possible unidentified parameters

** added covariates (step 1)

*** added covariates (final step)

Table 6 Average number of trips by type and their standard errors within each of the five clusters

	Cluster 1	s.e.	Cluster 2	s.e.	Cluster 3	s.e.	Cluster 4	s.e.	Cluster 5	s.e.
Cluster Size	0.3160	0.0028	0.3033	0.0031	0.143	0.0027	0.1296	0.0022	0.1080	0.0024
Trips Solo First Day	3.1745	0.0188	3.0992	0.0226	0.2726	0.0176	2.7029	0.0313	0.0064	0.0024
Trips With Relatives First Day	0.0183	0.0026	2.1351	0.0195	3.2324	0.0436	0.4594	0.0152	0.1221	0.0191
Trips With Others First Day	0.0113	0.0018	0.0757	0.0042	0.1072	0.0068	1.6744	0.023	0.16	0.0107
Trips with Unknown First Day	0.0005	0.0005	0.0014	0.0005	0.0005	0.0005	0.0162	0.0025	0.0067	0.0017
Trips Solo Second Day	3.0443	0.0184	2.9746	0.0221	0.191	0.0165	2.5826	0.0307	0.0029	0.0017
Trips With Relatives Second Day	0.0126	0.0024	2.2086	0.0196	3.1214	0.0441	0.4588	0.0149	0.1481	0.0169
Trips With Others Second Day	0.012	0.0017	0.0613	0.0038	0.1131	0.0066	1.7628	0.0239	0.1401	0.0102
Trips Unknown Second Day	0.0012	0.0008	0.0015	0.0004	0.0009	0.0005	0.0172	0.003	0.0058	0.0017

Table 7 Significance of person and household factors in cluster composition

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Employed at interview time (significant for all clusters)					
Yes	0.3838	0.3373	0.0680	0.1382	0.0727
No	0.1999	0.2450	0.2718	0.1147	0.1686
Age Group (* means not significant at 5%)					
15-17	0.0599	0.0519	0.3696	0.2954	0.2232
18-24	0.3018*	0.1730	0.1234	0.2596	0.1422*
25-34	0.3214	0.3015	0.1423	0.1295	0.1054*
35-44	0.2847	0.4278	0.1001	0.1124	0.0749
45-54	0.3940	0.3382	0.0695	0.1213	0.0770
55-64	0.3989	0.2784	0.1193*	0.1072	0.0963
65-98	0.2515*	0.2194	0.2721	0.0910	0.1660
Number of children ages 1 to 5 (significant for all clusters)					
0	0.3357	0.2853	0.1289	0.1375	0.1127
1	0.1840	0.4408	0.2106	0.0834	0.0813
2	0.1964	0.3891	0.2698	0.0704	0.0743
3	0.2531	0.2055	0.4343	0.0560	0.0511
4	0.1918	0.5020	0.2204	0.0858	0.0000
6	0.0008	0.5029	0.3488	0.0000	0.1474
Number of children ages 6 to 17 (significant for all clusters)					
0	0.3680	0.2447	0.1416	0.1286	0.1171
1	0.2338	0.4025	0.1308	0.1388	0.0941
2	0.1787	0.4684	0.1460	0.1280	0.0789
3	0.1459	0.4613	0.1866	0.1162	0.0899
4	0.0914	0.4183	0.2781	0.1198	0.0923
5	0.0801	0.5185	0.2015	0.0995	0.1004
6	0.0263	0.4218	0.3945	0.1504	0.0069

Note: Other variables included are: number of household cars, year of the survey, first year when the person entered panel, county of residence, and recruitment sample.

Table 8 The sequence of models estimated for activity and travel

		LogL	BIC	Parameters
Model 1	1-Cluster	-1185677	2371986	61
Model 2	2-Cluster	-1087068	2175254	108
Model 3	3-Cluster	-1055890	2113383	155
Model 4	4-Cluster	-1035485	2073060	202
Model 5	5-Cluster	-1029321	2061217	249

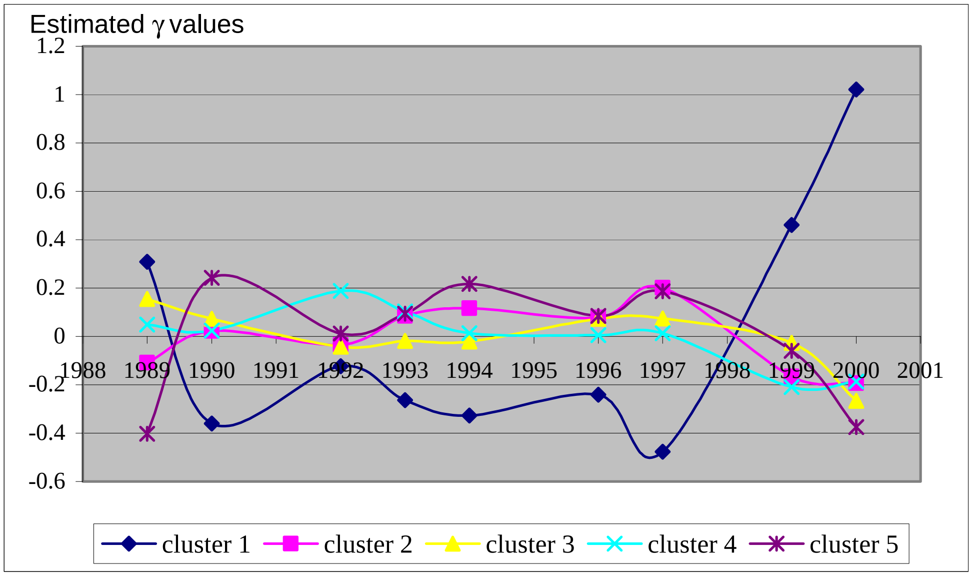
Table 9 Average values of criteria variables within each of the five activity-travel cluster

	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	
Variables used to create clusters (Y)						
	Average	**Average**	**Average**	**Average**	**Average**	**Total**
Trips in day 1	3.69	3.60	6.44	3.30	6.45	
Travel time to work related places in day 1	6.06	19.37	26.83	13.50	68.29	
Travel time to non work related places in day 1	62.53	40.23	80.89	44.23	107.72	
Total amount of time in all activities in day 1	283.89	541.70	595.32	498.02	473.70	
Trips in day 2	3.69	3.34	6.45	3.13	5.12	
Travel time to work related places in day 2	6.06	18.34	25.10	13.50	63.47	
Travel time to non work related places in day 2	62.22	38.92	97.15	43.86	70.49	
Total amount of time in all activities in day 2	277.71	531.42	602.20	498.82	451.21	
Cluster Size	0.318	0.2353	0.1889	0.1887	0.0691	1
Cluster Size in %	31.8	23.53	18.89	18.87	6.91	100
Number of cases	31117		BIC	2061217		
Number of parameters (Npar)	249		AIC	2059139		
Log-likelihood (LL)	-1029321		CAIC	2061466		

Table 10 Significance of person and household factors in activity-travel cluster composition

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Employed at interview time (significant for all clusters)					
Yes	0	0.3399	0.273	0.2912	0.0958
No	0.8636	0.0559	0.0446	0.0127	0.0232
Age Group					
15-17	0.7634	0.0873	0.0656	0.0765	0.0072
18-24	0.2699	0.2605	0.1884	0.2278	0.0533
25-34	0.1628	0.2956	0.2183	0.2477	0.0756
35-44	0.1342	0.305	0.248	0.2265	0.0863
45-54	0.1361	0.2952	0.236	0.2389	0.0937
55-64	0.3823	0.1973	0.1776	0.1743	0.0684
65-98	0.8105	0.0634	0.0572	0.0402	0.0287
Number of children ages 1 to 5 (significant for all clusters)					
0	0.3284	0.2307	0.1874	0.1857	0.0677
1	0.2401	0.2687	0.2068	0.2037	0.0806
2	0.2637	0.261	0.184	0.2144	0.0769
3	0.3434	0.2616	0.147	0.225	0.023
4	0.4615	0.0671	0.2313	0.2213	0.0187
6	0.5	0.0052	0.465	0	0.0298
Number of vehicles in household					
1-2	0.4183	0.2155	0.1643	0.1505	0.0513
3	0.3166	0.2411	0.1873	0.1868	0.0683
4	0.2784	0.2363	0.2038	0.2066	0.0748
5+	0.2368	0.2442	0.2062	0.2229	0.0898

Note: Other variables included are: year of the survey, first year when the person entered panel, county of residence, and recruitment sample.



Mean values of membership probabilities per cluster and year

Year of wave	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
1989	.3065	.3345	.1310	.1370	.0910
1990	.3102	.3207	.1370	.1412	.0910
1992	.3173	.3231	.1462	.1169	.0965
1993	.3131	.3068	.1429	.1223	.1150
1994	.3106	.2896	.1571	.1356	.1070
1996	.3086	.3145	.1415	.1202	.1153
1997	.3316	.2928	.1440	.1193	.1122
1999	.3030	.2842	.1607	.1388	.1133
2000	.3435	.2655	.1260	.1354	.1297
Total	.3160	.3033	.1430	.1296	.1080

Figure 1 Values of the γs and membership probabilities for each year in the panel for the joint-solo clusters

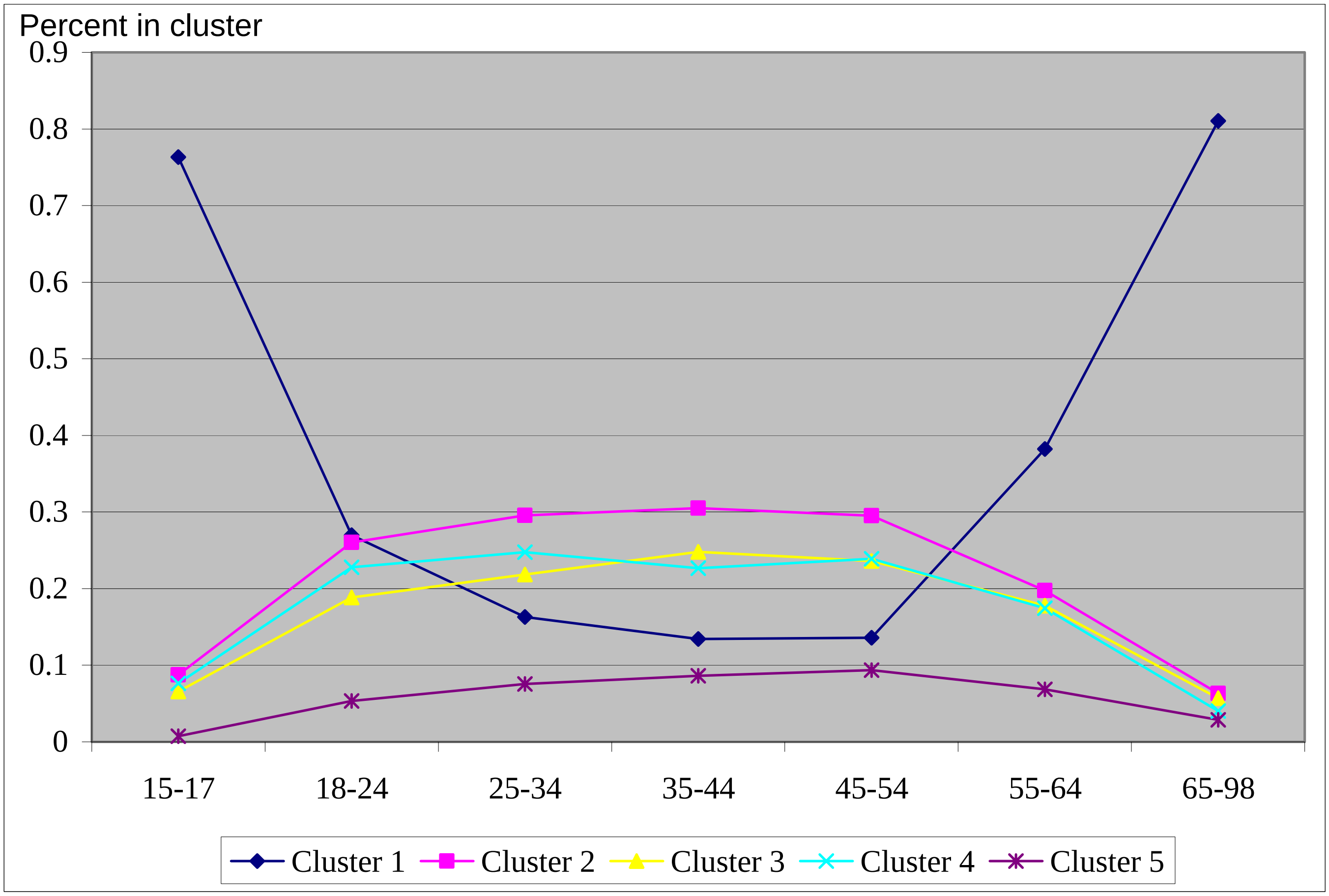


Figure 2 Average probability of cluster membership in relation to age group

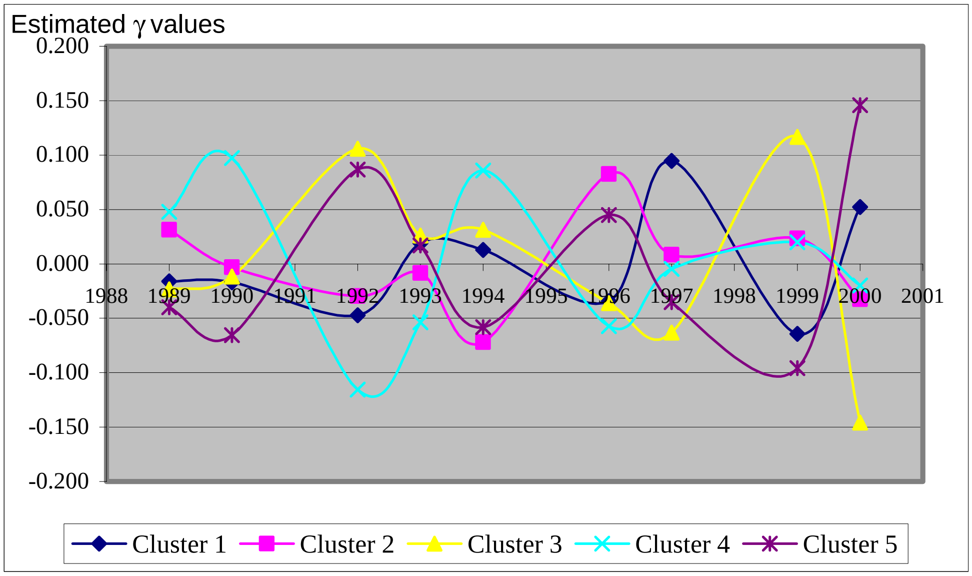


Figure 3 Values of the γs for year in the panel for the activity-travel clusters

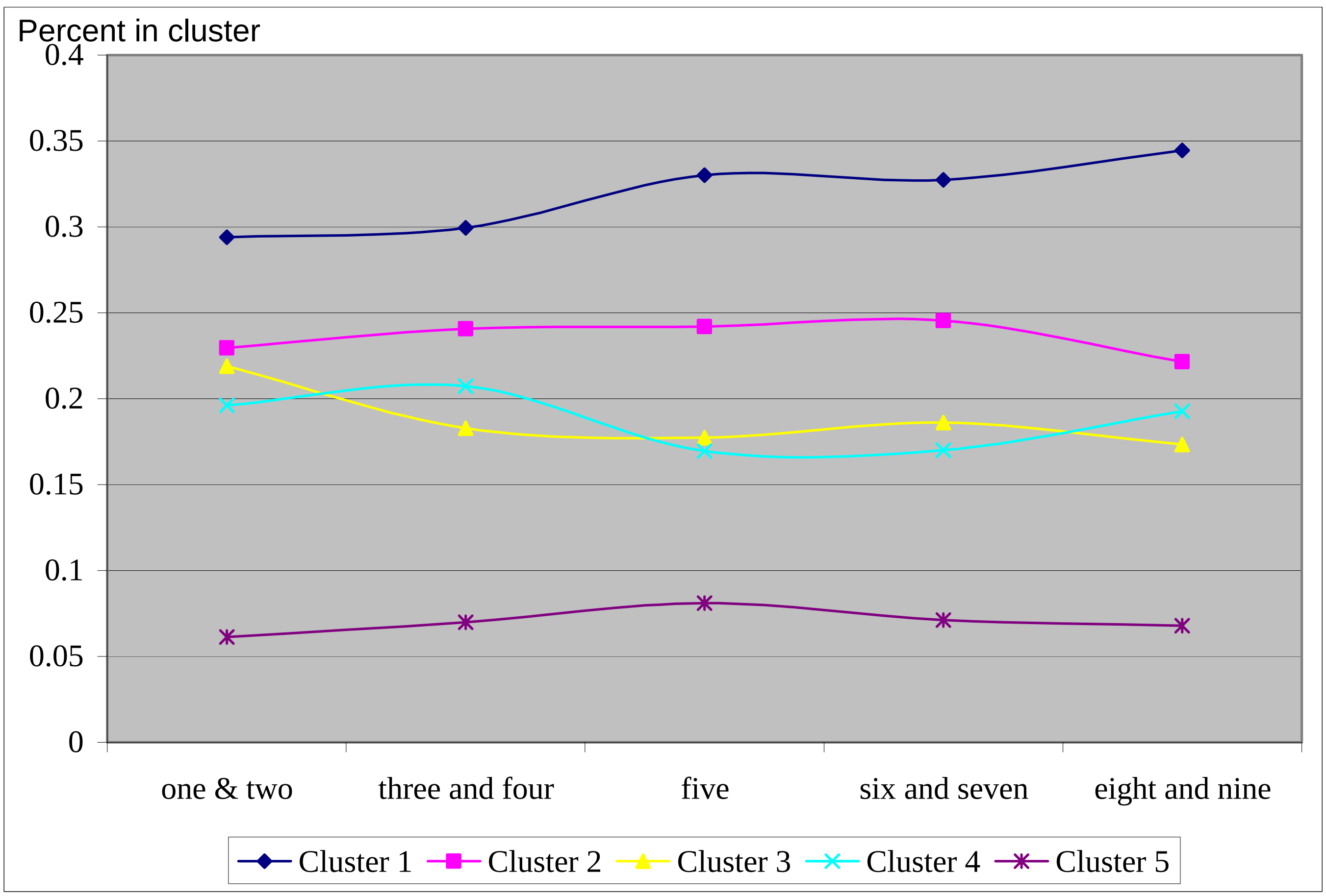


Figure 4 Time evolution of the activity and travel clusters in PSTP