PERGAMON

Transportation Research Part B 36 (2002) 59–82

TRANSPORTATION RESEARCH PART B

www.elsevier.com/locate/trb

Nonparametric identification of daily activity durations using kernel density estimators

Jeffrey P. Kharoufeh [a,1], Konstadinos G. Goulias [b,*]

[a] *Department of Industrial and Manufacturing Engineering, 310 Leonhard Building, The Pennsylvania State University, University Park, PA 16802, USA*

[b] *Department of Civil and Environmental Engineering and The Pennsylvania Transportation Institute, 201 Transportation Research Building, The Pennsylvania State University, University Park, PA 16802, USA*

Abstract

Modeling of activity duration has become an important aspect in characterizing activity and travel patterns. The standard approach for analyzing activity duration is to use hazard-based models to account for the duration dependence within an activity while estimating covariate effects and heterogeneity. The effects of covariates and heterogeneity on duration have been modeled using parametric and nonparametric regression techniques. In this paper, we present a nonparametric pattern recognition approach that can precede hazard-based duration modeling to identify heterogeneity patterns that may undermine duration modeling if not detected. The technique utilizes a kernel estimate of the probability density function (pdf) of various activity durations and allows for statistical compari son of distributions to evaluate differences between groups of individuals. Preliminary testing on the first wave of a panel survey indicates that the approach is insightful in evaluating covariate effects while allowing visual inspection of heterogeneity in the data. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords: Activity-based travel demand; Duration; Kernel density estimation

1. Introduction

Activity-based approaches to travel demand forecasting have been proposed as a new paradigm in travel behavior analysis because of their ability to model the derived nature of travel demand

* Corresponding author. Tel.: +1-814-863-7926; fax: +1-814-865-3039.

E-mail addresses: jpk175@psu.edu (J.P. Kharoufeh), kxg2@psu.edu (K.G. Goulias).

[1] Tel.: +1-814-865-9860; fax: +1-814-863-4745.

PII: S0191-2615(00)00038-2

(i.e., people travel in order to participate in various activities). The majority of these new frameworks advocate the use of behavioral simulation to create virtual, daily individual schedules (Barret et al., 1995; Ettema et al., 1995a,b; Kwan, 1995; Bowman and Ben-Akiva, 1996; Kitamura et al., 1996; Vaughn et al., 1997). To this end, significant advances have also been made in data collection specifically designed for these frameworks (Arentze et al., 1997). Methodological frameworks and models following this new paradigm in travel demand analysis and forecasting can be found in Jones (1990), Hamed and Mannering (1993), Ben-Akiva and Bowman (1995), Hofman et al. (1995), Morrison and Loose (1995), Pendyala et al. (1995), Bhat (1996a), Recker (1995), Golob (1996), Golob et al. (1996), Golob and McNally (1996), Miller (1996), Stopher and Lee-Gosselin (1996), Ettema and Timmermans (1997), Ma (1997).

The success of an activity-based forecasting system depends heavily on the behavioral models that depict and attempt to predict a person's daily schedule. However, facilitation of such analyses requires a variety of models (estimated from observed data and/or coded into a microsimulation model) that allocate time in activities and travel. Examples of variables of interest in such models are the time expenditure for each activity type, the frequency of activity types, the amount of time a person spends outside the home, and so forth. If models that explain each of these entities are available and can be embedded into a "simulator", then synthetic (or simulated-virtual) schedules can be produced. Hence, analyzing the activity participation behavior of individuals or households can provide meaningful insight to the prediction of travel demand.

Fundamental to activity-based forecasting systems are parametric or nonparametric regression models that reflect the relationships between covariates and the aforementioned dependent variables. Correct depiction of this relationship is based on the mathematical relationship between the covariates and the probability distribution of the dependent variable. Typically, the relationship between the dependent variable and its covariates has been assumed to be either linear or a form that may be linearized. Moreover, assumptions regarding the distribution of the dependent variable have spanned a wide spectrum of parametric distributions and a few ad-hoc nonparametric approaches. However, these approaches may have difficulty accounting for special features, such as bimodality, in the distributions. The current paper provides an illustration of a cross-classification alternative to regression models that can accommodate anomalies in the distribution of the dependent variable while making no assumptions about the functional relationship between the dependent variable and its explanatory variables (or covariates). The dependent variables in this study are the amount of time allocated to six activity types in a day by each segment of the population. The segments are defined for a person's lifecycle stage using a few explanatory variables such as age, presence of children in the household, and gender. It should be noted that the method here can be expanded to other dependent and explanatory variables used in the activity-based approaches including social and economic circumstances and the level of service offered by the urban setting and transportation system.

The next section reviews proportional hazard-based approaches and formally defines the problem. Section 3 gives a review of kernel density estimation, and in Section 4, we apply this nonparametric technique to duration modeling to investigate a few covariates and visually inspect heterogeneity. Section 5 presents the data used for evaluation of the technique and our empirical results. Section 6 summarizes our conclusions and future research in this area.

2. The problem

The standard approach for analyzing activity duration is to use parametric hazard-based models. This approach accounts for the fact that, in many situations, the time at which an activity will end is conditioned upon the amount of time already spent participating in the activity (i.e., duration dependence). The hazard function describes the rate at which we expect an activity to end given that an individual has been participating in the activity up to a certain time, t. Suppose a continuous random variable, T describing duration in an activity has probability density function (pdf), f and distribution function, F. The hazard function is defined as

$$\lambda(t) = f(t)(1 - F(t))^{-1}. \tag{1}$$

Characterization of the hazard function alone is not sufficient to specify activity patterns; it is also of interest to determine the effect of covariates on activity duration. The baseline hazard function (denoted by λ_0) is defined as a hazard function which does not account for these covariate effects. The proportional hazard function assumes that covariates act multiplicatively on the baseline hazard and is defined as (see Bhat, 1996a)

$$\lambda(t) = \lambda_0(t)\exp(-\beta'\mathbf{X} + \mathbf{w}), \tag{2}$$

where λ_0 is the baseline hazard function, $\mathbf{X}$ the covariate vector, β the vector of parameters to be estimated and $\mathbf{w}$ represents differences between individuals (unobserved heterogeneity). In the estimation of Eq. (2), it is necessary to specify a distribution for the baseline hazard function and the unobserved heterogeneity.

Parametric or nonparametric distributions may be used for the estimation of baseline duration distributions and the distribution of the unobserved differences between individuals. Most authors have used the parametric hazard-based approach for the modeling of activity duration. For instance, Paselk and Mannering (1994) modeled the delay of vehicles crossing the US and Canadian border in Washington. The authors point out that the most commonly used parametric distributions for the baseline hazard are the Weibull, exponential, lognormal and log-logistic and found that the log-logistic distribution best modeled the delay duration, although heterogeneity was not taken into account.

Other authors have taken a nonparametric approach to the problem. Most notably are Cox (1972) and Han and Hausman (1990). Cox uses a partial maximum likelihood approach, which is unable to easily account for unobserved heterogeneity and interval-level data. As noted by other authors (Kalbfleisch and Prentice, 1980; Bhat, 1996a), this approach is tedious in the presence of tied data. However, Mannering et al. (1994) have used the Cox approach for the modeling of home-stay duration. The Han and Hausman (1990) approach simultaneously estimates covariate effects and the parameters of the baseline hazard while accommodating tied data and unobserved heterogeneity. Bhat (1996a) has modeled the after-work shopping activity of individuals. His work entailed nonparametric estimation of both the baseline hazard and unobserved heterogeneity and concluded that in any case, a nonparametric specification of both distributions works best.

The objective of this work is to present experiments with a nonparametric approach for modeling activity duration that accounts for covariate effects by using a cross-classification alternative to the regression model of Eq. (2). The technique allows us to inspect differences between individuals by viewing the distinct shapes of nonparametric duration density functions. Although the density function, cumulative distribution function and the hazard function may be used interchangeably, the density function may reveal such features as bimodality in the activity durations that may be obscured when using the cumulative distribution or hazard functions. The density approximations are obtained using kernel estimates that reveal interesting behavioral relationships in a more elegant and parsimonious fashion than econometric techniques in the current literature. The authors are unaware of the application of kernel density estimates by researchers in the modeling of activity duration; however, kernel applications abound in the economics literature (see for example Bell and Pitt, 1998; Pritsker, 1998). It should be noted that the method illustrated here is not (as of yet) offered as an alternative modeling technique. Rather, it is used as a "pattern recognition" tool during the initial stages of data analysis. For example, the tool can be used to detect cross-sectional differences in time allocation. In addition, it can be used to study longitudinal differences when individuals move from one lifecycle stage to another while at the same time recognizing heterogeneity explicitly as shown later in the paper.

3. Overview of density estimation

Let X denote a continuous random variable and X_i, $i = 1, 2, \ldots, n$, denote n independent, identically distributed observations of X. Denote by f, the pdf for X. Density estimation is the attempt to either parametrically or nonparametrically approximate the pdf for X. In parametric density estimation, the focus is on fitting a theoretical probability distribution to the collected data. This approach requires the analyst to determine the appropriate theoretical distribution and its corresponding parameters. Nonparametric density estimation seeks to allow the data to stand on its own by using alternative methods to determine the form of the density function. The two major nonparametric approaches are the well-known histogram density estimator and kernel density estimator (KDE).

In his discussion of the use of parametric baseline hazard functions, Bhat (1996a) points out that there is often little theoretical justification for using any particular parametric distribution. Furthermore, certain distribution features, such as bimodality often cannot be modeled by a theoretical distribution; thus, the case for nonparametric density estimation is strengthened. In the following section, we review the well-known KDE, which requires specification of a single smoothing parameter that may be automatically calculated.

3.1. The kernel density estimator

This review of kernel estimators follows from Silverman (1986). The KDE is a completely nonparametric approach for estimating the density function of a continuous random variable and is best explained by first introducing a naïve estimator. Given a sample of n independent, identically distributed observations from a continuous, univariate distribution, $X_1, \ldots, X_n$, the pdf for the random variable, X may be defined as

$$f(x) = \lim_{h \to 0} \frac{\Pr\{x - h < X < x + h\}}{2h}. \tag{3}$$

Thus, an approximation for the density function given above may be obtained by allowing h to be very small and using the relative frequency definition for the numerator, yielding

$$\hat{f}(x) = N_x(2hn)^{-1}, \tag{4}$$

where N_x is the number of observations falling in the interval $(x - h, x + h]$. It is seen that if an arbitrary observation, X_i is within h-units of the fixed point of interest, x, then the observation adds an influence of $(2hn)^{-1}$ to the density estimate at x. In general, let

$$w(u) = \begin{cases} 0.5 & |u| < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Now, if an observation, X_i is within the h-radius of x, then the influence of such an observation is given by

$$\frac{1}{nh} w\left(\frac{x - X_i}{h}\right). \tag{6}$$

It is seen that if an observation contributes to the density estimate at x, then the estimator centers a box of width $2h$ and height $(2hn)^{-1}$ over the observation. Now by summing the "influence" of each observation, the naïve density estimate at x is given by

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} w\left(\frac{x - X_i}{h}\right). \tag{7}$$

Now replace the weight function of Eq. (7) with a different function, which places a kernel at each observation, rather than boxes. Such a function is referred to as the kernel function, K, thus defining the kernel estimate as

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right). \tag{8}$$

Here, K is usually (but not necessarily) a symmetric pdf itself, being positive valued and integrating to unity on the real line. The shape of the kernel function dictates the shape of each kernel mass while the bandwidth h determines its width. As the bandwidth tends to zero, spurious details appear in the distribution. On the other hand, as the bandwidth increases, the density estimate becomes smoother and less sensitive to the curvature of the true density.

In contrast to the histogram estimator which requires specification of the origin and bin width, the kernel density estimate requires only specification of the bandwidth, h. However, h may be chosen as a function of the kernel function, K, and the number of observations in the data sample. In the following sections, we discuss selection of the appropriate kernel function and bandwidth for kernel density estimation.

3.2. Selection of kernel function and bandwidth

The kernel function is generally a unimodal, symmetric pdf, such as the Gaussian kernel function. It is obvious by the definition equation (8) that if K is a pdf, then so too is $\hat{f}$. Silverman (1986) states that the choice of kernel function is not as significant a choice as the selection of the bandwidth. Thus, the choice of kernel function may be based upon other considerations such as computational effort and desired mathematical properties (e.g., differentiability). We use the Gaussian kernel because it is a unimodal, symmetric density and allows for automatic bandwidth selection. Due to the importance of appropriate bandwidth selection, we shall next focus on this aspect of kernel density estimation and its relation to the sample size, n.

Intuitively, one is interested in selecting a smoothing parameter that is optimal in some sense. Hence, it is natural to think in terms of minimizing some measure of discrepancy between the estimated density function and its true form. The most widely accepted global measure of discrepancy was first proposed by Rosenblatt (1956) in which he defined the mean integrated squared error (MISE) as

$$\mathrm{MISE}(\hat{f}) = E\int_{-\infty}^{\infty}\left\{\hat{f}(x) - f(x)\right\}^2 \mathrm{d}x, \tag{9}$$

where f is the unknown true density. With some simplification, Eq. (9) decomposes into the sum of the integrated variance and integrated bias terms. Assuming that K is a symmetric density with zero mean and nonzero variance, approximate expressions for the variance and bias may be obtained (see Silverman, 1986). Interestingly, the bias does not directly depend on the sample size (n) but does depend on the bandwidth (h), which is proportional to $n^{-1/5}$. The optimal smoothing parameter for minimizing the approximate MISE is obtained by choosing the appropriate K subject to certain conditions. Consequently, the Epanechnikov kernel (Epanechnikov, 1969) is the function that minimizes the approximate MISE, and thus, all efficiencies of symmetric kernels are reported with respect to the Epanechnikov. Silverman (1986, p. 43) gives the efficiencies of some commonly used kernels. In this work, the Gaussian kernel (with efficiency 0.9512) was used and is given by

$$K_{\mathrm{G}}(t) = (2\pi)^{-1/2}\exp\{-t^2/2\}, \quad -\infty \leqslant t \leqslant \infty \tag{10}$$

and its optimal smoothing parameter may be computed directly by

$$h_{\mathrm{opt}} = 0.9\min\{\sigma,\ \text{interquartile range}/1.34\}n^{-1/5}, \tag{11}$$

where σ may be approximated with the sample standard deviation. Silverman (1986) states that this value works well for a wide range of univariate distributions, including those with a bimodal structure, making it suitable for our application to activity duration distributions. Furthermore, the author demonstrates through simulation that for $n = 100$, skewness or bimodality will be revealed using the smoothing parameter of Eq. (11). One drawback of the kernel estimator is its tendency to show spurious noise in long-tailed distributions. This is due to the fact that a fixed smoothing parameter is used over the range of the data. However, by employing adaptive kernel methods, which vary the smoothing parameter over the range of the data, this problem can be alleviated.

4. Model of activity duration

Consider a general setting in which we assume that duration data (i.e., time expenditure in a particular activity) are available for various daily activities of individuals. Furthermore, we assume that the data are segmented into a finite number of disjoint sets representing various lifecycle stages. It is well known that lifecycle stages influence travel behavior. In addition, when studying the dynamics of travel behavior, lifecycle stage and transitions from one stage to another may capture a significant portion of behavioral variation over time.

Let I denote the finite set of daily activities in which individuals participate and J denote the finite set of lifecycle groups. In this preliminary work, the strong assumption of independence between different activities in I will be employed. However, Bhat (1996b) presents a competing risk hazard model that allows for dependence when heterogeneity is included. Define T_{ij} as the continuous random variable denoting the amount of time individuals in the group $j \in J$ spend participating in activity $i \in I$. We seek to estimate the probability distribution of T_{ij}. Denote by $\hat{f}_{ij}$, the kernel density estimate of the true pdf for T_{ij}. The density estimate is given by

$$\hat{f}_{ij}(t) = \frac{1}{n_{ij}h_{ij}} \sum_{k=1}^{n_{ij}} K\left(\frac{t - T_{ijk}}{h_{ij}}\right), \tag{12}$$

where n_{ij} is the number of independent, identically distributed observations for T_{ij}, h_{ij} the optimal smoothing parameter obtained by Eq. (11) and T_{ijk} is the kth duration observation for an individual in group j participating in activity i.

Although activity duration times are naturally continuous random variables, we choose to estimate their distributions by discrete approximations. The rationale for doing so is as follows. The primary objective of density estimation in this work is the visualization and exploration of the data. Hence, practical considerations require estimation of the density over a finite grid of points to achieve this visualization. Since most activity duration times in our study are reported in multiples of 5 min, it seems intuitive to approximate the density on such a grid. These 5-min tendencies in the reported data were also observed by Bhat (1996a) in shopping activity duration data. Let $Q_{ij} \equiv \{t : t = 5k,\ k = 0, 1, \ldots, M\}$ be a finite support for the distribution of T_{ij}, where M is a positive integer chosen large enough so as to contain the sample space of the random variable. Using the kernel density estimates by Eq. (12), a probability mass may be calculated for each $t \in Q_{ij}$. Let $p_{ij}(t)$ denote the probability mass at t given by

$$p_{ij}(t) = \Pr\{T_{ij} = t\} = \delta^{-1}\hat{f}_{ij}(t), \tag{13}$$

where $\delta = \sum_{t \in Q_{ij}} \hat{f}_{ij}(t)$ normalizes the density so that the masses sum to unity over the finite support. In essence, Eq. (13) constructs a histogram estimator of the duration times in which each cell width is equal to zero. Jones (1989) demonstrates that even when the kernel estimate is discretized, the appropriate optimal smoothing parameter is still proportional to $n^{-1/5}$. Thus, the effect of discretization is minimal when we use the appropriate smoothing parameter. As mentioned above, the Gaussian kernel given by Eq. (10) is used throughout this work because it is a unimodal, symmetric distribution, which allows for automatic selection of the smoothing parameter.

5. Data and empirical results

Data used in this study for the purpose of evaluating the nonparametric approach originated from the Puget Sound Transportation Panel (Goulias and Ma, 1996). This database is comprised of results of a survey conducted in the Seattle metropolitan area in the fall of 1989. The data consist of five waves with each wave containing a two-day travel survey. Each survey contains information regarding household and person demographics and socio-economic characteristics, reported travel behavior, and level of service offered by the transportation system. For each individual in this wave, variables were created for subsistence, maintenance, leisure (in-home), leisure (out-of-home), home-stay and travel duration in minutes (see definitions in Section 5.1). Participants in the first wave (year) of the panel that continued for at least three additional waves (1621 individuals) were used. In the first stage of testing, the data from all lifecycle groups for each day were aggregated and day-to-day variation for each activity type was examined. In the second stage, the data were partitioned into 10 lifecycle groups to investigate a few covariate effects.

The rationale for using the lifecycle as an explanatory factor of travel behavior is as follows. Individuals in households of different composition (in terms of number of adults and children, their age and gender composition, and their labor force participation status) are expected to behave differently (Jones et al., 1983). Moreover, as households and their members move from one lifecycle stage to another (e.g., from a two-adult household to a two-adult and a child), they are subject to noticeable changes in their time allocation and travel behavior. Detecting and understanding these changes are of paramount importance for longitudinal analysis (Kitamura and Kostyniuk, 1986; Goodwin, 1997; van Beek et al., 1997), and key to realistic social and demographic microsimulation models for travel demand forecasting (Goulias, 1992; Chung, 1997; Goulias and Kitamura, 1997).

However, due to societal changes, the definition of lifecycle stages may need to constantly be updated. For example, older stage definitions, such as nuclear family, single adult and surviving senior member, may be less capable of differentiating behavior. To accommodate and incorporate recent social trends, these definitions of lifecycle stages have been modified as in Chicoine and Boyle (1984) and Kostyniuk et al. (1989), emphasizing the single-adult household. Lifecycle stages in past studies have been defined ad-hoc based on social theories and trends, study objectives, and data availability (Zimmerman, 1982; Kostyniuk and Kitamura, 1982; McDonald and Stopher, 1983; Kitamura and Kostyniuk, 1986; Kostyniuk et al., 1989; Goulias, 1992; Vadarevu and Stopher, 1996; Stopher and Metcalf, 1999). In a few studies, statistical techniques have also been used to derive lifecycle stages (e.g., Chung, 1997; Marker, 2000). The findings in these past studies, with a few pragmatic considerations, are the basis for our lifecycle stage sample segmentation herein.

The following key considerations motivate our definition of the lifecycle stages used in this paper: (a) within each stage, the age distribution in the household captures differences in the propensity to participate in activities and travel; (b) gender is (still) a fundamental parameter of differentiation across different task and time allocations in households; (c) the presence of children and their age distribution within a household are fundamental motivators for differences in time allocation across household members; (d) transition out of the labor force is a fundamental factor in explaining differences in time allocation; and (e) the kernel estimator used here requires approximately 40–50 observations in each segment.

Table 1 shows the variables and grouping included in the study in addition to the number of individuals (n) in each group. Note that in this way, we account for differences in gender, age, and presence of children in the household.

Table 1
Summary of data segmentation

Group no.	Life-cycle stage	n
1	Male parent of young child (<6 yr)	132
2	Female parent of young child (<6 yr)	136
3	Male parent of older child (6–17 yr)	154
4	Female parent of older child (6–17 yr)	178
5	Single male (all age groups)	45
6	Single female (all age groups)	102
7	Male, double income, no children (18–64 yr)	302
8	Female, double income, no children (18–64 yr)	323
9	Senior male (65 yr or older)	118
10	Senior female (65 yr or older)	131

Segmentation of the data into 10 mutually exclusive and exhaustive groups allows for a pairwise comparison of the estimated duration densities by means of goodness-of-fit tests. The availability of a probability distribution for each of the groups provides a glimpse of the "big picture". That is, the distribution contains much more information about the behavior of the groups than summary statistics alone (e.g., mean, variance, kurtosis, etc.). In addition to the qualitative assessment of covariate effects and heterogeneity attained through the shape of the distributions, the goodness-of-fit test provides a quantitative measure to confirm (or refute) differences between the groups.

The Kolmogorov–Smirnov (KS) two-sample test is used to determine if two independent samples have been drawn from the same population. A discussion of the KS two-sample test can be found in any good text on nonparametric statistics; this discussion follows from Conover (1980). When the two-tailed test is employed, it is sensitive to any differences between the distributions and the populations from which they originated. Furthermore, the test is exact when the random variables considered are continuous. Let F and G be the true distribution functions for two samples. The KS two-sample test performs the hypothesis test (see Conover, 1980)

$$\begin{aligned} &H_0 : F(x) = G(x) \quad -\infty \leqslant x \leqslant \infty, \\ &H_1 : F(x) \neq G(x) \quad \text{for at least one } x \text{ on the real line.} \end{aligned} \tag{14}$$

The two-tailed test statistic for the KS test is given by

$$T = \sup_{x \in Q} |S_1(x) - S_2(x)|, \tag{15}$$

where $S_1(x)$ and $S_2(x)$ are the empirical distribution functions for the two independent samples and Q is the set of points at which the distribution functions are evaluated. This is a natural measure of the difference in "vertical" distance between the two distribution functions. The decision rule is to reject the null hypothesis at the α level of significance if T exceeds the $1 - \alpha$ quantile of the distribution. If the number of observations for both samples is fixed at n, then the critical value at the 0.05 level for large samples sizes ($n > 40$) is given by

$$T^* = \frac{1.92}{\sqrt{n}}. \tag{16}$$

Thus, if the KS test is not significant, then the two samples are assumed to originate from the same population. That is, we expect the two groups under consideration to exhibit the same behavior in terms of daily time of participation in a particular activity.

5.1. Model with no covariate effects (aggregate data)

Initially, the data were not segmented and the nonparametric approach of Section 4 was used to test for the day-to-day variation for each of the six activity duration variables: subsistence, maintenance, leisure (in-home), leisure (out-of-home), home-stay, and travel. Activity engagement information, and the daily activity participation indicators used here, have been extracted from the travel diary. To make the analysis tractable, the original out-of-home activity types, equivalent to trip purposes, are aggregated into subsistence (work, college, school), maintenance (shopping, appointment, personal), out-of-home leisure (visiting, free time), and the activities at-home are divided into those that are for leisure (in-home leisure) and those for all other activities (home-stay). Travel is the total time spent traveling for any reason during a day.

Table 2 gives the KS two-sample test statistic (T) and the corresponding critical value (T^*).

Fig. 1 shows a comparison of the distributions for subsistence duration on two consecutive weekdays (the mean duration for subsistence activity 279.27 min in day 1 and 272.53 min in day 2). Throughout this work, distribution functions were compared at 186 points

Table 2
Summary of day-to-day variation[a]

Activity	T	T^*
Subsistence	0.0143	0.1408
Maintenance	0.0245	0.1408
Leisure (in-home)	0.0127	0.1408
Leisure (out-of-home)	0.0070	0.1408
Home-stay	0.0070	0.1408
Travel	0.0216	0.1408

[a] Note: All tests performed at the 0.05 level with $n = 186$.

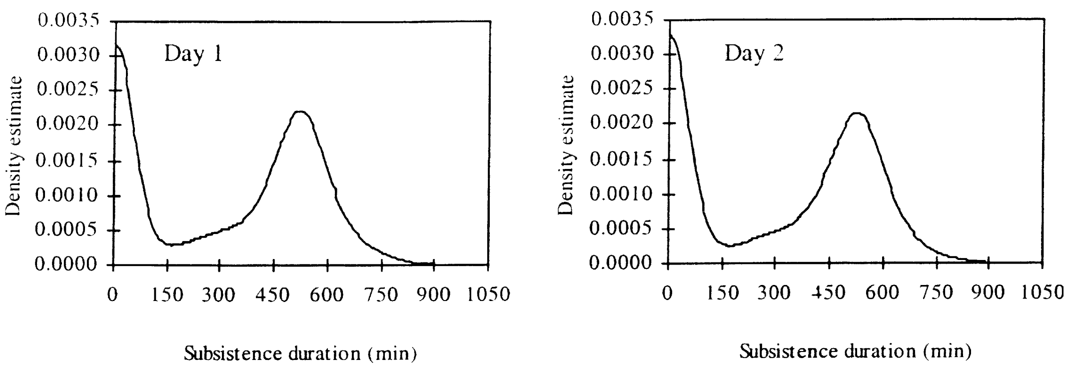

Fig. 1. Day-to-day variation in aggregate data for subsistence duration.

$(t = 0, 5, 10, \ldots, 925)$ for which the true distribution is approximated. It is seen that for all cases, there is no statistically significant difference in the time individuals spent participating in a particular activity from one weekday to the next when the data are not separated by lifecycle group. This agrees with past analyses (Ma and Goulias, 1997) in which time homogeneity was shown by tests applied to aggregates of people, reconfirming that data aggregation leads to apparent homogeneity. The day-to-day tests presented later will provide a clearer picture regarding this apparent "stability" in travel behavior over time.

5.2. Model with covariate effects (segmentation)

The second stage of experimentation involved segmentation of the 1621 observations into 10 distinct lifecycle stages. This segmentation allows us to investigate the effect of a few covariates such as gender, age, marital status and the presence and age of children in the household. Within an activity on a given day, distributions were compared pairwise to determine if significant differences exist between groups. Tables 3–8 (included in Appendix A) summarize these results for each of the activities and each of the two days in the first wave of the data. In each of these tables, the entry gives the calculated KS test statistic when testing the hypothesis (at the 0.05 level) that the distribution of individuals in the row group is equal to the distribution of individuals in the column group for a given activity and day. Key observations in these tables are as follows:

(a) *Senior males and females*: As expected, senior males and females have significantly different subsistence activity distributions than all other groups in both days. This is clearly due to their exit from the labor force. It is also interesting to note the significant difference in subsistence activity between males and females, a difference is observed in maintenance activity as well. However, as we move to other activity types, differences with other lifecycle groups begin to disappear. This is particularly evident in travel time duration in which senior females have identical distributions with other groups of people and senior males have different distributions in the second day only.

(b) *Parents of young children*: Male and female parents of young children have different subsistence, maintenance, out-of-home leisure and in-home leisure durations. This may be an indication of a clear task allocation in which females rear the children while males pursue other activities. Females in this group have significantly different activity durations from all other groups except female parents of older children for maintenance and in-home leisure. Males of young children are not significantly different than their counterpart male parents of older children.

(c) *Parents of older children*: As mentioned above, the behavior of male parents of older children is similar to that of male parents of young children; however, we also find this behavior similar to that of single males. Female parents of older children are significantly different than their male counterparts.

(d) *Persons in double income household*: Lack of children in the household and the presence of two working persons indicate practically indistinguishable activity distributions for males and females.

(e) *Single persons*: In many comparisons, single persons have duration distributions similar to the "double income no children" group. However, within this group, males and females differ in their subsistence, out-of-home leisure, and home-stay duration during the second day.

Table 3
KS test statistic for subsistence duration in days 1 and 2[a]

	Male parent of young child	Female parent of young child	Male parent of older child	Female parent of older child	Single male	Single female	Male, double income, no children	Female, double income, no children	Senior male	Senior female
Subsistence duration (*day* 1)										
Male parent of young child	–	0.4887	0.0704	0.2998	0.1209	0.3003	0.1128	0.2169	0.7570	0.6422
Female parent of young child	–	–	0.4639	0.2674	0.3821	0.2048	0.3768	0.2902	0.3937	0.2242
Male parent of older child	–	–	–	0.2582	0.0885	0.2706	0.0884	0.1864	0.7287	0.6143
Female parent of older child	–	–	–	–	0.1881	0.0643	0.2050	0.0906	0.5964	0.4640
Single male	–	–	–	–	–	0.1823	0.0248	0.1029	0.6838	0.5603
Single female	–	–	–	–	–	–	0.1903	0.0854	0.5575	0.4106
Male, double income, no children	–	–	–	–	–	–	–	0.1145	0.6613	0.5450
Female, double income, no children	–	–	–	–	–	–	–	–	0.5928	0.4700
Senior male	–	–	–	–	–	–	–	–	–	0.1890
Senior female	–	–	–	–	–	–	–	–	–	–
Subsistence duration (*day* 2)										
Male parent of young child	–	0.5000	0.0878	0.3176	0.1210	0.2891	0.1374	0.259	0.7093	0.6360
Female parent of young child	–	–	0.4686	0.2405	0.4270	0.2206	0.3752	0.2746	0.3660	0.2283
Male parent of older child	–	–	–	0.2579	0.0516	0.2483	0.0938	0.1996	0.6965	0.6138
Female parent of older child	–	–	–	–	0.2068	0.0532	0.1822	0.0628	0.5634	0.4494
Single male	–	–	–	–	–	0.2091	0.0549	0.1527	0.6795	0.5838
Single female	–	–	–	–	–	–	0.1551	0.0621	0.5348	0.4146
Male, double income, no children	–	–	–	–	–	–	–	0.1213	0.6255	0.5291
Female, double income, no children	–	–	–	–	–	–	–	–	0.5510	0.4473
Senior male	–	–	–	–	–	–	–	–	–	0.1507
Senior female	–	–	–	–	–	–	–	–	–	–

[a] Note: All tests performed at the 0.05 level with $n = 186$, critical value $= 0.1408$.

Table 4
KS test statistic for maintenance duration in days 1 and 2[a]

	Male parent of young child	Female parent of young child	Male parent of older child	Female parent of older child	Single male	Single female	Male, double income, no children	Female, double income, no children	Senior male	Senior female
Maintenance duration (*day* 1)										
Male parent of young child	–	0.3963	0.0818	0.3172	0.1391	0.3967	0.1061	0.2113	0.4167	0.4269
Female parent of young child	–	–	0.3340	0.0938	0.3011	0.0528	0.2960	0.2066	0.0442	0.0655
Male parent of older child	–	–	–	0.2507	0.0627	0.3350	0.0392	0.1378	0.3545	0.3672
Female parent of older child	–	–	–	–	0.2130	0.1119	0.2160	0.1167	0.1172	0.1371
Single male	–	–	–	–	–	0.3038	0.0832	0.0964	0.3230	0.3360
Single female		–	–	–	–	–	0.2964	0.2123	0.0286	0.0333
Male, double income, no children	–	–	–	–	–	–	–	0.1078	0.3164	0.2380
Female, double income, no children	–	–	–	–	–	–	–	–	0.2295	0.2456
Senior male	–	–	–	–	–	–	–	–	–	0.0304
Senior female	–	–	–	–	–	–	–	–	–	–
Maintenance duration (*day* 2)										
Male parent of young child	–	0.3695	0.0386	0.29693	0.1205	0.2806	0.0420	0.1333	0.3265	0.1874
Female parent of young child	–	–	0.3421	0.0793	0.4588	0.0990	0.3281	0.2773	0.0495	0.2011
Male parent of older child	–	–	–	0.2719	0.1577	0.2528	0.0417	0.1054	0.2991	0.1595
Female parent of older child	–	–	–	–	0.3796	0.0429	0.2579	0.1721	0.0563	0.1229
Single male	–	–	–	–	–	0.3612	0.1578	0.2250	0.4093	0.2631
Single female	–	–	–	–	–	–	0.2386	0.1523	0.0501	0.1024
Male, double income, no children	–	–	–	–	–	–	–	0.0917	0.2851	0.1454
Female, double income, no children	–	–	–	–	–	–	–	–	0.2006	0.0541
Senior male	–	–	–	–	–	–	–	–	–	0.1523
Senior female	–	–	–	–	–	–	–	–	–	–

[a] Note: All tests performed at the 0.05 level with $n = 186$, critical value = 0.1408.

Table 5
KS test statistic for leisure (in-home) duration in days 1 and 2[a]

	Male parent of young child	Female parent of young child	Male parent of older child	Female parent of older child	Single male	Single female	Male, double income, no children	Female, double income, no children	Senior male	Senior female
Leisure (*in-home*) (*day* 1)										
Male parent of young child	–	0.2331	0.1143	0.2105	0.1343	0.0457	0.0720	0.0251	0.1760	0.1471
Female parent of young child	–	–	0.1934	0.0567	0.1231	0.2375	0.1836	0.2111	0.0603	0.1209
Male parent of older child	–	–	–	0.1396	0.0795	0.1145	0.0773	0.0938	0.1682	0.1530
Female parent of older child	–	–	–	–	0.0934	0.2078	0.1582	0.1892	0.0375	0.0919
Single male	–	–	–	–	–	0.1351	0.0813	0.1140	0.1301	0.1140
Single female	–	–	–	–	–	–	0.1116	0.0563	0.2032	0.1864
Male, double income, no children	–	–	–	–	–	–	–	0.0591	0.1241	0.0766
Female, double income, no children	–	–	–	–	–	–	–	–	0.1547	0.1340
Senior male	–	–	–	–	–	–	–	–	–	0.0607
Senior female	–	–	–	–	–	–	–	–	–	–
Leisure (*in-home*) (*day* 2)										
Male parent of young child	–	0.1540	0.0605	0.0821	0.0737	0.1519	0.1020	0.0590	0.0452	0.1565
Female parent of young child	–	–	02101	0.1210	0.2109	0.2771	0.2272	0.2006	0.1796	0.2813
Male parent of older child	–	–	–	0.0988	0.0453	0.1431	0.0934	0.0497	0.1000	0.1479
Female parent of older child	–	–	–	–	0.1172	0.2325	0.1788	0.1345	0.1269	0.2346
Single male	–	–	–	–	–	0.1352	0.0904	0.0478	0.0862	0.1445
Single female	–	–	–	–	–	–	0.0549	0.1003	0.1069	0.0518
Male, double income, no children	–	–	–	–	–	–	–	0.0519	0.0650	0.0580
Female, double income, no children	–	–	–	–	–	–	–	–	0.0824	0.1054
Senior male	–	–	–	–	–	–	–	–	–	0.1131
Senior female	–	–	–	–	–	–	–	–	–	–

[a] Note: All tests performed at the 0.05 level with $n = 186$, critical value $= 0.1408$.

Table 6
KS test statistic for leisure (out-of-home) duration in days 1 and 2[a]

	Male parent of young child	Female parent of young child	Male parent of older child	Female parent of older child	Single male	Single female	Male, double income, no children	Female, double income, no children	Senior male	Senior female
Leisure (*out-of-home*) (*day* 1)										
Male parent of young child	–	0.1419	0.1127	0.2129	0.1850	0.0936	0.1335	0.1479	0.2620	0.1850
Female parent of young child	–	–	0.0667	0.0760	0.1040	0.1716	0.0616	0.0426	0.1617	0.1293
Male parent of older child	–	–	–	0.1175	0.1506	0.1424	0.0832	0.0901	0.2106	0.1807
Female parent of older child	–	–	–	–	0.0795	0.2426	0.1375	0.0894	0.0935	0.0942
Single male	–	–	–	–	–	0.2066	0.0747	0.0915	0.1069	0.0328
Single female	–	–	–	–	–	–	0.1120	0.1585	0.2942	0.2147
Male, double income, no children	–	–	–	–	–	–	–	0.0481	0.1891	0.1071
Female, double income, no children	–	–	–	–	–	–	–	–	0.1410	0.1127
Senior male	–	–	–	–	–	–	–	–	–	0.0868
Senior female	–	–	–	–	–	–	–	–	–	–
Leisure (*out-of-home*) (*day* 2)										
Male parent of young child	–	0.2617	0.0666	0.0853	0.1253	0.0647	0.0808	0.1912	0.1797	0.1852
Female parent of young child	–	–	0.2139	0.1829	0.1514	0.3043	0.1812	0.0794	0.0940	0.0920
Male parent of older child	–	–	–	0.0451	0.1163	0.1120	0.1057	0.1700	0.1240	0.1268
Female parent of older child	–	–	–	–	0.0807	0.1297	0.0678	0.1322	0.0973	0.1027
Single male	–	–	–	–	–	0.1695	0.0601	0.0719	0.0767	0.0748
Single female	–	–	–	–	–	–	0.1251	0.2355	0.2240	0.2295
Male, double income, no children	–	–	–	–	–	–	–	0.1104	0.0990	0.1044
Female, double income, no children	–	–	–	–	–	–	–	–	0.0642	0.0589
Senior male	–	–	–	–	–	–	–	–	–	0.0244
Senior female	–	–	–	–	–	–	–	–	–	–

[a] Note: All tests performed at the 0.05 level with $n = 186$, critical value = 0.1408.

Table 7
KS test statistic for home-stay duration in days 1 and 2[a]

	Male parent of young child	Female parent of young child	Male parent of older child	Female parent of older child	Single male	Single female	Male, double income, no children	Female, double income, no children	Senior male	Senior female
Home-stay duration (*day* 1)										
Male parent of young child	–	0.3254	0.1081	0.2189	0.0551	0.0722	0.0860	0.1646	0.0846	0.1342
Female parent of young child	–	–	0.2311	0.1501	0.3112	0.3729	0.3695	0.4218	0.2544	0.4101
Male parent of older child	–	–	–	0.1151	0.0883	0.1766	0.1847	0.2581	0.0591	0.2329
Female parent of older child	–	–	–	–	0.2021	0.2790	0.2824	0.3448	0.1401	0.3281
Single male	–	–	–	–	–	0.0973	0.1110	0.1876	0.0887	0.1592
Single female	–	–	–	–	–	–	0.0796	0.0985	0.1545	0.1020
Male, double income, no children	–	–	–	–	–	–	–	0.0801	0.1654	0.0482
Female, double income, no children	–	–	–	–	–	–	–	–	0.2389	0.0740
Senior male	–	–	–	–	–	–	–	–	–	0.2136
Senior female	–	–	–	–	–	–	–	–	–	–
Home-stay duration (*day* 2)										
Male parent of young child	–	0.1968	0.0603	0.1557	0.1462	0.3382	0.2115	0.2242	0.1581	0.1724
Female parent of young child	–	–	0.1787	0.1384	0.3096	0.4729	0.3711	0.3814	0.3431	0.3503
Male parent of older child	–	–	–	0.1018	0.1652	0.3847	0.2635	0.2763	0.2184	0.2019
Female parent of older child	–	–	–	–	0.2027	0.4509	0.3419	0.3539	0.3065	0.2988
Single male	–	–	–	–	–	0.3701	0.2490	0.2617	0.2016	0.1683
Single female	–	–	–	–	–	–	0.1417	0.1302	0.2137	0.3020
Male, double income, no children	–	–	–	–	–	–	–	0.0363	0.0896	0.1614
Female, double income, no children	–	–	–	–	–	–	–	–	0.1228	0.1741
Senior male	–	–	–	–	–	–	–	–	–	0.1005
Senior female	–	–	–	–	–	–	–	–	–	–

[a] Note: All tests performed at the 0.05 level with $n = 186$, critical value = 0.1408.

Table 8
KS test statistic for travel duration in days 1 and 2[a]

	Male parent of young child	Female parent of young child	Male parent of older child	Female parent of older child	Single male	Single female	Male, double income, no children	Female, double income, no children	Senior male	Senior female
Travel duration (*day* 1)										
Male parent of young child	–	0.0851	0.0841	0.0714	0.1332	0.1170	0.0221	0.1105	0.1447	0.1355
Female parent of young child	–	–	0.1074	0.0479	0.0725	0.0458	0.0808	0.0349	0.0803	0.0791
Male parent of older child	–	–	–	0.0715	0.1688	0.1163	0.0669	0.1231	0.1402	0.1311
Female parent of older child	–	–	–	–	0.1189	0.0481	0.0533	0.0713	0.0844	0.0859
Single male	–	–	–	–	–	0.0807	0.1410	0.0607	0.1081	0.1111
Single female	–	–	–	–	–	–	0.0981	0.0301	0.0392	0.0423
Male, double income, no children	–	–	–	–	–	–	–	0.0991	0.1253	0.1224
Female, double income, no children	–	–	–	–	–	–	–	–	0.0486	0.0515
Senior male	–	–	–	–	–	–	–	–	–	0.0300
Senior female	–	–	–	–	–	–	–	–	–	–
Travel duration (*day* 2)										
Male parent of young child	–	0.1085	0.0323	0.0620	0.1115	0.1193	0.0472	0.1098	0.1531	0.2854
Female parent of young child	–	–	0.1354	0.0497	0.0324	0.0705	0.0799	0.0161	0.0762	0.1927
Male parent of older child	–	–	–	0.0857	0.1342	0.1212	0.0786	0.1327	0.1592	0.2944
Female parent of older child	–	–	–	–	0.0501	0.0736	0.0397	0.0479	0.1051	0.2283
Single male	–	–	–	–	–	0.0496	0.0852	0.0225	0.0732	0.1881
Single female	–	–	–	–	–	–	0.0869	0.0547	0.0910	0.1826
Male, double income, no children	–	–	–	–	–	–	–	0.0868	0.1410	0.2677
Female, double income, no children	–	–	–	–	–	–	–	–	0.0612	0.1817
Senior male	–	–	–	–	–	–	–	–	–	0.1375
Senior female	–	–	–	–	–	–	–	–	–	–

[a] Note: All tests performed at the 0.05 level with $n = 186$, critical value = 0.1408.

Although the presence of children in the household has a major impact on female activity distributions, this factor does not affect the counterpart male activity distributions. In addition, large sample variation across groups is due to females who have different activity densities across

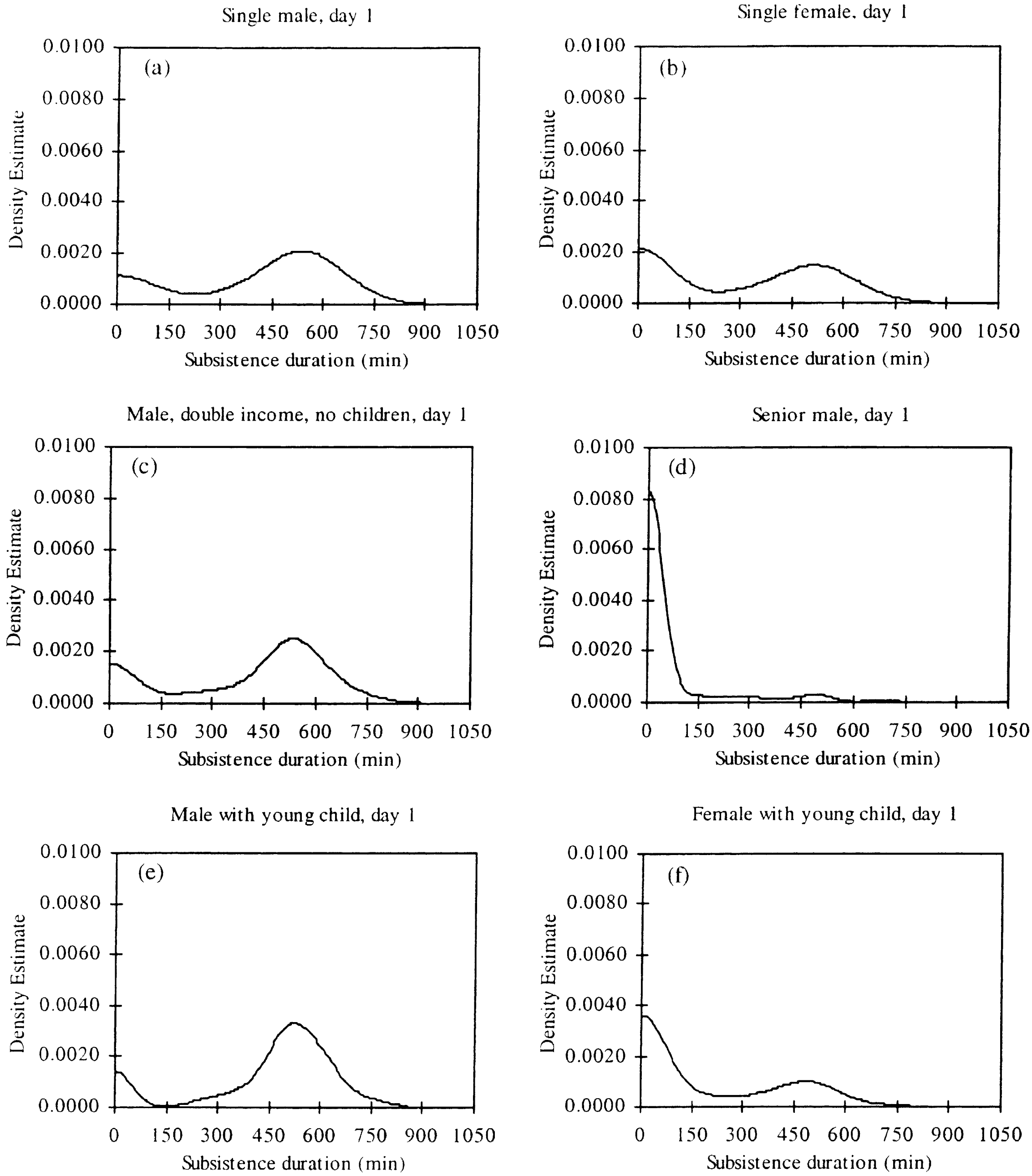

Fig. 2. Sample subsistence density estimates for various groups.

lifecycle stages. Males have similar behavior across lifecycle stages until they become seniors and become inactive. However, these are differences and similarities across groups of people at a given point in time. Evidence on evolutionary aspects (change of behavior moving from lifecycle stage

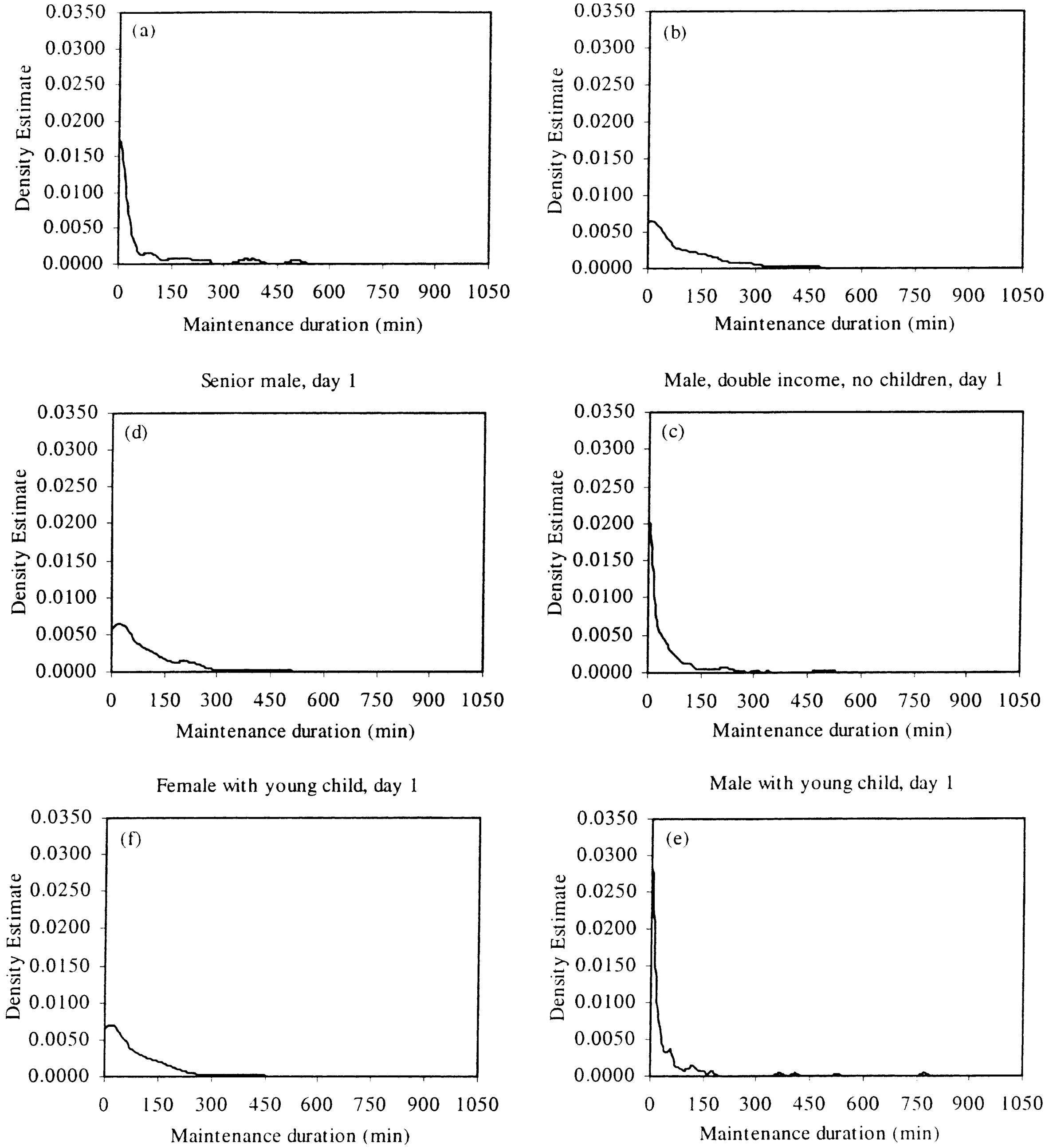

Fig. 3. Sample maintenance density estimates for various groups.

Table 9
KS test statistic for day-to-day variation by activity type[a]

	Subsistence	Maintenance	Leisure (in-home)	Leisure (out-of-home)	Home-stay	Travel
Male parent of young child	0.0262	0.0803	0.1209	0.0668	0.1132	0.0693
Female parent of young child	0.0212	0.0352	0.0702	0.1319	0.0393	0.0544
Male parent of older child	0.0121	0.0484	0.0599	0.0668	0.0866	0.0483
Female parent of older child	0.0302	0.0625	0.0211	0.1268	0.0552	0.0362
Single male	0.0360	0.1631	0.0639	0.1336	0.1386	0.0919
Single female	0.0279	0.0763	0.1114	0.0692	0.1899	0.0507
Male, double income, no children	0.0194	0.0246	0.0417	0.0282	0.0567	0.0325
Female, double income, no children	0.0271	0.0271	0.0456	0.0960	0.0389	0.0409
Senior male	0.0361	0.0615	0.1077	0.1372	0.1329	0.0299
Senior female	0.0154	0.2061	0.1590	0.0802	0.1835	0.1365

[a] Note: All tests performed at the 0.05 level, $n = 186$, critical value = 0.1408.

to lifecycle stage) needs to be confirmed by analyzing subsequent panel waves. Figs. 2 and 3 provide sample pdfs generated by the kernel method for subsistence and maintenance duration, respectively.

Finally, day-to-day variation within groups was analyzed for each activity in order to discern if the aggregate results hold up when we consider individual groups. Table 9 gives a summary of all results.

Surprisingly, given the substantial day-to-day variation in activity and travel behavior observed in previous analyses (Ma and Goulias, 1997) using different behavioral indicators, Table 9 shows very little change in the activity distributions from one weekday to the next. All activity types for all groups, with the exception of single males and senior females maintenance, single females and senior females for home-stay, and senior females for in-home leisure, are statistically the same between the two days. Using similar behavioral variables with lifecycle definitions derived using a statistical method, Marker (2000) reaches conclusions similar to those in this paper. Recalling that the method used here does not impose any strong assumptions on the data structure, we are tempted to believe that past indications may be partially due to model construction, inadequate treatment of heterogeneity, and differences in the behavioral variables used. However, a more comprehensive and comparative study using the kernel technique should be conducted. This can be accomplished in many ways. For example, one can repeat the analysis for the remaining waves of the panel to verify if individuals, as they move from one stage to another, change their time allocation significantly. In addition, one can also extend the method to other behavioral indicators and perform comparisons with other studies (e.g., activity and travel counts of episodes as in Ma and Goulias, 1999).

6. Conclusions and future research

This paper presents experiments using a nonparametric pattern recognition tool for the purpose of investigating covariate effects and heterogeneity. The technique utilizes a kernel estimate of the pdf of various activity durations and allows for statistical comparison of distribution functions to evaluate differences between groups of individuals. Within this framework, it is not necessary to make restrictive assumptions about the distribution of activity duration; nor is it necessary to assume a particular mathematical relationship between covariates and the dependent variable (duration). Moreover, the kernel method is unaffected by potential bimodality in distributions, offering another option when estimating baseline hazards. The kernel technique was tested on six activity durations using the first wave of a panel survey and contributes insight into behavioral relationships that is unavailable with the use of summary statistics alone.

The empirical analysis demonstrates the potential for using this technique as an alternative to regression models, as several behaviorally interesting results are obtained. First, the data indicate a statistically significant difference between senior males and females and all other groups for subsistence and maintenance activities. This finding is clearly due to their exit from the labor force. Second, significant differences in most activity distributions are noted between males and females when they are the parents of young children. Third, the data indicate the presence of children in the household having a major impact on female activity distributions, while males appear to be unaffected by the presence of children. Fourth, large variation across groups is seen between female groups while males have similar behavior across lifecycle groups until they become seniors. In spite of these substantial cross-sectional differences across groups, all activity types for all groups (with the exception of a few cases) are remarkably similar between the two days analyzed. These results, however, need to be verified exploring other travel behavior indicators and the remaining waves of the panel. The technique illustrated here cannot be considered a viable alternative to proportional hazard-based duration models until further testing is complete and a comparison with competing methods is made.

This work identifies many potential areas of future research in the method and its application to duration modeling. First, it may be instructive to consider the use of adaptive kernel methods when a bimodal nature is observed. If the variances of the two modes differ dramatically, a fixed bandwidth may oversmooth the density estimate. Adaptive kernel methods vary the bandwidth over the range of the data in an attempt to account for these differences. Second, the density estimates presented herein can potentially be used within the framework of a proportional hazard model, allowing for a completely distribution-free estimate of the baseline hazard and unobserved heterogeneity distribution (see Gonzalez-Monteiga et al., 1996 for kernel-based hazard models). The method could also be instructive in exploring dependencies between different activities by applying the multivariate version of the kernel method. A logical extension is to next apply the technique to different waves in the panel in order to analyze the year-to-year daily activity duration evolution over time.

Acknowledgements

The authors wish to acknowledge the suggestions made by anonymous referees that substantially improved the paper. This research was sponsored by the Mid-Atlantic Universities Transportation Center (MAUTC), Region III, US Department of Transportation and the Center for Intelligent Transportation Systems (CITranS) at the College of Engineering, The

Pennsylvania State University. This research was also made possible by funds from the Weiss Graduate Fellowship for the first author and the Fulbright/FLAD Chair award for the second author.

Appendix A

Kolmogorov–Smirnov statistics for data segmentation experiment (Tables 3–8).

References

Arentze, T., Hofman, F., Kalfs, N., Timmermans, H., 1997. Data needs, data collection and data quality requirements of activity-based transport demand models. Presented at Transport Surveys, Raising the Standard, Grainau, Germany, May 24–30.

Barret, C., Berkbigler, K., Smith, L., Loose, V., Beckman, B., Davis, J., Roberts, D., Williams, M., 1995. An operational description of TRANSIMS. LA-UR-95-2393, Los Alamos, NM.

Bell, B.D., Pitt, M.K., 1998. Trade union decline and the distribution of wages in the UK: evidence from kernel density estimation. Oxford Bull. Econom. Stat. 60, 509–528.

Ben-Akiva, M.E., Bowman, J.L., 1995. Activity-based disaggregate travel demand model system with daily activity schedules. Presented at Activity-Based Approaches: Activity Scheduling and the Analysis of Activity Patterns, Eindhoven, The Netherlands, 25–29 May.

Bhat, C., 1996a. A hazard-based duration model of shopping activity with nonparametric baseline specification and nonparametric control for unobserved heterogeneity. Transport. Res. B 30, 189–207.

Bhat, C., 1996b. A generalized multiple durations proportional hazard model with an application to activity behavior during the evening work-to-home commute. Transport. Res. B 30, 465–480.

Bowman, J.L., Ben-Akiva, M., 1996. Activity based travel forecasting. Tutorial transcript from the TMIP Conference on Activity-Based Travel Forecasting, New Orleans, LA.

Chicoine, J.E., Boyle, D.K., 1984. Lifecycle concept: a practical application to transportation planning. Transport. Res. Rec. 987, 1–7.

Chung, J., 1997. Transportation impact simulation of access management and other land use policies using geographic information systems, Ph.D. Thesis, Penn State University, USA.

Conover, W.J., 1980. Practical Nonparametric Statistics. Wiley, New York.

Cox, J., 1972. Regression models and life tables. J. R. Statist. Soc. B 26, 186–220.

Epanechnikov, V.A., 1969. Nonparametric estimation of a multidimensional probability density. Theor. Probab. Appl. 14, 153–158.

Ettema, D., Borgers, A., Timmermans, H., 1995a. A competing risk-hazard model of activity choice, timing, sequencing, and duration. Transport. Res. Rec. 1493, 101–109.

Ettema, D., Borgers, A., Timmermans, H., 1995b. Simulation Model of Activity Scheduling Heuristics (SMASH): empirical test and simulation issues. Presented at the Activity Based Approaches: Activity Scheduling and the Analysis of Activity Patterns. Eindhoven, The Netherlands, 25–29 May.

Ettema, D.F., Timmermans, H. (Eds.), 1997. Activity-Based Approaches to Travel Analysis. Pergamon Elsevier, The Netherlands.

Golob, T.F., 1996. A model of household demand for activity participation and mobility. UCI-ITS-WP-96-5. Institute of Transportation Studies, University of California at Irvine, Irvine, CA.

Golob, T.F., Bradley, M.A., Polak, J., 1996. Travel and activity participation as influenced by car availability and use. Presented at the 75th Annual Meeting of the Transportation Research Board, Washington, DC.

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

Gonzalez-Monteiga, W., Cao, R., Marron, J.S., 1996. Bootstrap selection of the smoothing parameter in nonparametric hazard rate estimation. J. Am. Stat. Assoc. 91 (435), 1130–1140.

Goodwin, P.B., 1997. Have panel surveys told us anything new?. In: Golob, T.F., Kitamura, R., Long, L. (Eds.), Panels for Transportation Planning Methods and Applications. Kluwer Academic Publisher, Boston, MA.

Goulias, K.G., 1992. Forecasting the impact of sociodemographic changes on travel demand: experiments with a dynamic microsimulation model system. Working Paper No. 94, The University of California Transportation Center.

Goulias, K.G., Ma, J., 1996. Analysis of longitudinal data from the Puget Sound transportation panel. Task B: Integration of PSTP Databases and PSTP Codebook. Final Report No. 9619. The Pennsylvania Transportation Institute, Penn State University, University Park, PA.

Goulias, K.G., Kitamura, R., 1997. Regional travel demand forecasting with dynamic microsimulation models. In: Panels for Transportation Planning: Methods and Applications. Kluwer, Dordrecht.

Hamed, M., Mannering, F.L., 1993. Modelling travelers' postwork activity involvement: toward a new methodology. Transport. Sci. 27, 381–394.

Han, A., Hausman, J.A., 1990. Flexible parametric estimation of duration and competing risk models. J. Appl. Econometrics 5, 1–28.

Hofman, F., Borgers, A.W.J., Timmermans, H., 1995. The necessity of activity-based modelling. Presented at Activity-Based Approaches: Activity Scheduling and the Analysis of Activity Patterns, Eindhoven, The Netherlands, 25–29 May.

Jones, M.C., 1989. Discretized and interpolated kernel density estimates. J. Am. Stat. Assoc. 84 (407), 733–741.

Jones, P.M., Dix, M.C., Clarke, M.I., Heggie, I.G., 1983. Understanding Travel Behaviour. Gower, Aldershot, England.

Jones, P.M., 1990. Developments in dynamic and activity-based approaches to travel analysis. A compendium of papers from the 1989 Oxford Conference, Avebury, UK.

Kalbfleisch, J.D., Prentice, R.L., 1980. The Statistical Analysis of Failure Time Data. Wiley, New York.

Kitamura, R., Kostyniuk, L.P., 1986. Maturing motorization and household travel: the case of nuclear-family households. Transport. Res. 20 (3), 245–260.

Kitamura, R., Pas, E.I., Lula, C.V., Lawton, T.K., Benson, P.E., 1996. The sequenced activity mobility simulator (SAMS): An integrated approach to modeling transportation, land use and air quality. Transportation 23, 267–291.

Kostyniuk, L.P., Kitamura, R., 1982. Lifecycle and household time–space paths: empirical investigation. Transport. Res. Rec. 879, 28–37.

Kostyniuk, L.P., Kitamura, R., Goulias, K.G., 1989. Mobility of single parents: what do the trio records show?. Special. Transport. Plann. Practice 3 (3), 203–217.

Kwan, M., 1995. GISICAS: An activity-based spatial decision support system for ATIS. Presented at Activity-Based Approaches: Activity Scheduling and the Analysis of Activity Patterns, Eindhoven, The Netherlands, 25–29 May.

Ma, J., 1997. An activity-based and micro-simulated travel forecasting system: a pragmatic, synthetic scheduling approach. Ph.D. Thesis, Penn State University, USA.

Ma, J., Goulias, K.G., 1997. A dynamic analysis of person and household activity and travel patterns using data from the first two waves in the Puget Sound Transportation Panel. Transportation 24, 309–331.

Ma, J., Goulias, K.G., 1999. Application of Poisson regression models to activity frequency analysis and prediction. Transport. Res. Rec. 1676, 86–94.

Mannering, F., Murakami, E., Kim, S.G., 1994. Temporal stability of travelers' activity choice and home-stay duration: some empirical evidence. Transportation 21, 371–392.

Marker, J.T., 2000. Period effects and cohort effects in lifecycles: preliminary analysis. Transportation Research Record. Presented at the 79th Annual Meeting of the Transportation Research Board, 9–13 January 2000, Washington, DC.

McDonald, K.G., Stopher, P.R., 1983. Some contrary indications for the use of household structure in trip-generation analysis. Transport. Res. Rec. 944, 92–100.

Miller, E.J., 1996. Microsimulation and activity-based forecasting. Presented at TMIP Conference on Activity-Based Travel Forecasting, New Orleans, LA.

Morrison, J., Loose, V., 1995. TRANSIMS model design criteria as derived from federal legislation. LAUR 95-1909. Los Alamos National Laboratory.

Paselk, T., Mannering, F., 1994. Use of duration models for predicting vehicular delay at a US Canadian border crossing. Transportation 21, 249–270.

Pendyala, R., Kitamura, R., Reddy, D.V.G.P., 1995. A rule-based activity-travel scheduling algorithm integrating neural networks of behavioral adaptation. Presented at Activity-Based Approaches: Activity Scheduling and the Analysis of Activity Patterns, Eindhoven, The Netherlands, 25–29 May.

Pritsker, M., 1998. Nonparametric density estimation and tests of continuous-time interest rate models. Rev. Financial Studies 11 (3), 449–487.

Recker, W.W., 1995. The household activity pattern problem: general formulation and solution. Transport. Res. 29, 61–77.

Rosenblatt, M., 1956. Remarks on some nonparametric estimates of a density function. Ann. Math. Statist. 42, 1815–1842.

Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. Chapman and Hall, London.

Stopher, P.R., Lee-Gosselin, M. (Eds.), 1996. Understanding Travel Behaviour in an Era of Change, Pergamon, Elsevier Science, Oxford, UK.

Stopher, P.R., Metcalf, H.M.A., 1999. Household activities, lifecycle and role allocation: tests on data sets from Boston and Salt Lake City, Paper 00472 in Preprint CD ROM, 78th Annual Meeting of the Transportation Research Board, 10–14 January, Washington, DC.

Vadarevu, R., Stopher, P.R., 1996. Household activities, lifecycle and role allocation. Transport. Res. Rec. 1556, 77–85.

van Beek, P., Kalfs, N., Blom, U., 1997. Gender differences in activities and mobility in The Netherlands, 1975–1990. Transport. Res. Rec. 1607, 134–138.

Vaughn, K.M., Speckman, P., Pas, E.I., 1997. Generating household activity-travel patterns (HATPs) for synthetic populations. Presented at the 76th Annual Meeting of the Transportation Research Board, Washington, DC.

Zimmerman, C.A., 1982. The lifecycle concept as a tool for travel research. Transportation 11, 51–69.