

## **INTERSHOPPING DURATION: AN ANALYSIS USING MULTIWEEK DATA**

Chandra R. Bhat, Teresa Frusti, Huimin Zhao,  
Stefan Schönfelder, and Kay W. Axhausen

### **ABSTRACT**

This study examines the rhythms in the shopping activity participation of individuals over a multiweek period by modeling the duration between successive shopping participations. A hazard based duration model is used to model intershopping duration, and a latent segmentation method is applied to distinguish between erratic shoppers and regular shoppers. The paper applies the methodology to examine the regularity and frequency of shopping behavior of individuals using a continuous six-week travel survey collected in the cities of Halle and Karlsruhe in Germany in the fall of 1999. The empirical results underscore the need to adopt a flexible hazard model form for analyzing intershopping durations. The results also provide important insights into the determinants of the regularity and frequency of individuals' shopping activity participation behavior.

Keywords: Multiday analysis, activity-travel behavior, intershopping duration, latent segmentation, hazard-based duration model, unobserved heterogeneity.

## 1. INTRODUCTION

The generation of the number of out-of-home activity episodes (or stops) of individuals is an important component of an activity-based analysis framework that emphasizes travel as being derived from the need to participate in activities (see Bhat and Koppelman, 1999 or Pendyala and Goulias, 2002 for recent comprehensive reviews of the activity-based travel analysis approach). Several earlier activity analysis studies have focused on activity stop generation, either in isolation or jointly with other stop attributes such as location, duration, sequencing, and travel time to stop (for recent examples, see Wen and Koppelman, 1999; Misra and Bhat, 2000; Bhat and Singh, 2000; Pendyala *et al.*, 2002; Bowman and Ben-Akiva, 2000; Kitamura and Fujii, 1998; Arentze and Timmermans, 2002).

The studies of stop generation identified above, and most other earlier studies in the activity analysis field, have used a single day as the basis of analysis. Unfortunately, such single day analyses implicitly assume uniformity in activity decisions from one day to the next, and do not allow the examination of variability in behavior over longer periods of time. In addition, single day analyses do not recognize that individuals who have quite dissimilar patterns on the survey day may in fact be similar in their patterns over a longer period of time. Such a case would arise if, for example, two individuals have the same behavioral pattern over a week, except that their cyclic patterns are staggered. Similarly, single day analyses do not recognize that individuals who appear similar in their patterns on the survey day may have very different patterns over longer periods of time. The net result is that models based on a single day of survey may reflect arbitrary statistical correlations, rather than capturing underlying behavioral relationships. Consequently, models based on a single day of analysis may be unsuitable for the analysis of transportation policy actions, as discussed by Jones and Clark (1988). Specifically,

Jones and Clark emphasize that multiday data is essential to extract information about the distribution of participation over time. The distribution of participation, in turn, provides important information regarding the frequency of exposure of different sociodemographic and travel segments to policy scenarios. For example, when examining the impact of land use mixing policies that encourage activity chaining, and/or congestion pricing policies, on shopping trips, it is important to know whether an individual participates in shopping activity everyday or whether the individual has a weekly shopping rhythm.

The focus of this paper is on activity stop generation within the larger context of a multiday activity generation model system. As indicated earlier, several previous studies have developed a conceptual and modeling framework for activity-based policy analysis within a single day framework. These frameworks, which include activity stop generation as an important component, can be extended to a multiday setting with a multiday activity stop generation module. The current effort contributes to the development of such a multiday activity stop generation module. In the next section, we briefly review earlier multiday studies of activity and travel behavior. In Section 1.2, we position the current research in the context of previous research.

### **1.1. Literature Review of Multiday Studies**

Earlier multiday studies of activity-travel behavior may be classified into three broad groups, as discussed in the next three paragraphs.

The first group of multiday studies uses descriptive analysis techniques to measure the extent of day-to-day variability in activity and travel characteristics (day-to-day variability refers to variations across days in activity and travel characteristics). Examples of such studies include

Pas and Sundar (1995) and Muthyalagari *et al.*, (2001). Pas and Sundar examine day-to-day variability in several travel indicators using a three-day travel diary data collected in 1989 in Seattle, while Muthyalagari *et al.* study intrapersonal variability using GPS-based travel data collected over a period of six days in Lexington, Kentucky. The latter study found larger day-to-day variability in travel indicators compared to the former, suggesting that GPS-based data collection may be recording short and infrequent trips better than traditional travel diary surveys.

The second group of multiday studies examines both the extent of day-to-day variability in activity-travel patterns as well as the influence of individual characteristics on the extent of variability. Most of the multiday studies fall in this category. Pas and Koppelman (1987) and Pas (1988) examine intrapersonal variability in daily number of trips using a seven-day activity data collection in 1973 in Reading, England. Pas and Koppelman (1987) develop a set of hypotheses about the impact of sociodemographic variables on intrapersonal variability, and test these hypotheses by comparing the amount of intrapersonal variability across predefined sociodemographic segments. Pas (1988), on the other hand, first clusters multiday activity travel patterns into a relatively small number of classes, and then examines the sociodemographic characteristics that distinguish the clusters. Pas's approach is a multiday extension of the methodology developed earlier by Pas (1983) and Koppelman and Pas (1984) to classify daily activity-travel patterns. Hanson and Huff (1986; 1988a; 1988b) and Huff and Hanson (1986; 1990) also examine day-to-day variability, with a focus on identifying relatively homogenous sociodemographic groupings based on observed multiday activity-travel behavior. Their studies, based on a multiweek travel survey conducted in Uppsala, Sweden in 1971, indicate that the amount of variability in behavior is intricately related to the complexity or detail used to represent activity-travel patterns. Their results also suggest that survey periods of longer than a

week may be needed to capture the distinct activity-travel behavior rhythms exhibited by individuals. In a more recent study of work commuting behavior, Mahmassani (1997) descriptively examine the effect of commuter characteristics and the commuter's travel environment on the likelihood of changing departure time and route choice from one day to the next for the morning home-to-work trip. Hatcher and Mahmassani (1992) focus on the same travel dimensions as Mahmassani (1997), except that their emphasis is on the evening work-to-home commute rather than the morning home-to-work commute. A ten-day diary data of morning and evening commute characteristics collected in Austin in 1989 is used in both these studies. Finally, Schlich (2001) has recently used a sequence alignment method to analyze intrapersonal variability in travel behavior using a 6-week travel survey conducted in Germany in the fall of 1999.

The third group of multiday studies uses multiday data to accommodate unobserved heterogeneity across individuals in models of activity-travel behavior (unobserved heterogeneity refers to differences among individuals in their activity-travel choices because of unobserved individual-specific characteristics). The objective of this group of studies is to recognize interpersonal variability in activity-travel behavior due to unobserved factors, and to distinguish this interpersonal variability from intrapersonal variability. While the end objective of this group of studies and the earlier two groups is a separation of interpersonal and intrapersonal variability, there is a subtle motivational difference. Studies of unobserved heterogeneity originate from a desire to control for differences in habitual and trait factors across individuals (*i.e.*, interpersonal variability), while the earlier two groups of studies are motivated from a desire to recognize within-individual differences in behavior (*i.e.*, intrapersonal variability). Of course, intrapersonal and interpersonal variability are simply two sides of the same total variability "coin". Examples

of studies focusing on unobserved individual heterogeneity include Bhat (2000a) and Bhat (1999). Bhat (2000a) examines unobserved heterogeneity in the context of work commute mode choice, while Bhat (1999) studies unobserved heterogeneity in the context of the number of non-work commute stops made by commuters. A multiday travel survey data collected in the San Francisco Bay area in 1990 is used in both studies.

## **1.2 The Current Research in the Context of Earlier Research**

The above studies have contributed substantially to our understanding of multiday travel behavior. The studies by Pas and his colleagues, Hanson and Huff, Muthyalagari *et al.*, and Schlich have quantified the magnitude of intrapersonal and interpersonal day-to-day variability in activity-travel behavior, and identified sociodemographic and locational attributes that impact this variability. A limitation of these studies, however, is that they do not explicitly disentangle the two quite different sources of day-to-day variability: (1) variability due to different choices made across days for regular daily decisions (for example, choosing different travel modes for the work trip), and (2) variability due to the non-daily nature of activity decisions (for instance, grocery shopping stops are not likely to be made every day). The studies by Mahmassani and Bhat, on the other hand, have focused only on the first source of variability, since these studies examine variability only in regular daily commuting patterns.

In contrast to earlier research that has either not explicitly disentangled the two different sources of day-to-day variability or focused only on variability due to different choices for regular daily decisions, the current research focuses on variability due to the non-daily nature of activity decisions. More precisely, the focus of the current study is on examining the rhythms in the shopping activity participation of individuals over a multiweek period (the reader will note

that it is the rhythms in activity participation over extended periods of time that are responsible for the day-to-day variability associated with non-daily activity decisions). Within the context of shopping activity, the current study focuses on maintenance-related shopping (including grocery shopping and medical drug shopping). In the rest of the paper, we will use the term “shopping” to refer to “maintenance-related shopping” for ease in presentation.

A continuous six-week travel survey collected in the cities of Halle and Karlsruhe in Germany in the fall of 1999 is used in the empirical analysis. The rhythms in shopping activity participation are examined by modeling the duration between successive shopping activity participations of individuals. The intershopping duration is measured in days, since a vast majority of individuals have no more than a single shopping activity participation on any given day. The methodology uses a hazard-based duration model structure since such a structure recognizes the dynamics of intershopping duration; that is, it recognizes that the likelihood of participating in shopping activity depends on the length of elapsed time since the previous participation. The hazard duration formulation also allows different individuals to have different rhythms in behavior and is able to predict shopping activity participation behavior over any period of time (such as a day, a week, or a month).

## **2. APPLICATION OF HAZARD MODELS TO INTERSHOPPING DURATION ANALYSIS AND METHODOLOGICAL CONTRIBUTION OF PAPER**

Hazard models have seen substantial use in the biometrics and economics fields, and are seeing increasing use in the transportation field (see Hensher and Mannering, 1994 and Bhat, 2000b for an extensive discussion of hazard-based duration models and transportation-related applications). In the context of intershopping durations, there have been two recent applications of hazard models, one by Schönfelder and Axhausen (2000) and the other by Kim and Park (1997).

Schönfelder and Axhausen examine the periodicity in intershopping durations using the same data source as the one used in the current study. Their study provides useful insights into the determinants of intershopping duration. However, it uses a Weibull parametric approach for the intershopping duration distribution or the Cox partial likelihood estimation approach. A potential problem with the parametric approach is that it inconsistently estimates the baseline hazard and the covariate effects when the assumed parametric form is incorrect (Meyer, 1990). Similarly, there are several limitations of the Cox approach. First, the dynamics of duration is of direct interest in studying the rhythms in shopping activity participation; the Cox approach, however, conditions out the parameters corresponding to the dynamics of duration. Second, the Cox approach becomes cumbersome in the presence of many tied failure times (Kalbfleisch and Prentice, 1980, page 101). As we will note later, tied failure times are the norm in intershopping durations. Third, unobservable heterogeneity (*i.e.*, variations across individuals in the intershopping duration due to unobserved individual factors) cannot be accommodated within the Cox partial likelihood framework without the presence of multiple integrals of the same order as the number of observations (see Han and Hausman, 1990). In addition to the issues discussed above, the paper by Schönfelder and Axhausen does not differentiate between individuals who have regularly spaced intershopping durations (regular shoppers) and individuals who do not have regularly spaced intershopping durations (erratic shoppers).

Kim and Park (1997) differentiate between regular and erratic shoppers by treating the shopper's trip regularity as a latent variable. However, their study does not include explanatory variables and it uses a parametric hazard form. Further, the classification of individuals as regular or erratic is based on posterior segment membership probabilities, which requires

information on the shopping activity history of individuals. Such information will not be available for individuals outside the estimation sample.

In the current paper, we develop a formulation that (a) accommodates a non-parametric baseline hazard, (b) endogenously classifies individuals as erratic or regular shoppers based on their demographic and household location characteristics, (c) includes the effect of relevant sociodemographic variables on intershopping duration, and (d) accommodates unobserved heterogeneity across individuals in intershopping durations. In addition, our approach recognizes the interval-level nature of intershopping durations; that is, it recognizes that a day is an interval of time, with several individuals having the same intershopping duration. The parametric and Cox approaches used by Schönfelder and Axhausen, and the formulation used by Kim and Park, employ density function terms in their respective likelihood functions that are appropriate only for estimation from continuous duration data.

From a methodological standpoint, the current paper extends the hazard-based formulation of Han and Hausman (1990) and Bhat (1996) to include a latent segmentation scheme to classify individuals into regular and erratic shoppers. The latent segmentation approach has been used in discrete choice modeling (see Bhat, 1997 and the many references therein), but, to our knowledge, this is the first application of a latent segmentation scheme for duration modeling that accommodates the effect of explanatory variables on the propensity to belong to each segment and on intershopping duration within each segment. The formulation in this paper may also be considered an extension of the latent segmentation procedures used commonly in discrete choice models; specifically, earlier latent segmentation studies in the context of discrete choice models have assumed a homogenous model relationship across

individuals within each segment, while the current study allows unobserved individual heterogeneity within each segment.

### **3. MODEL STRUCTURE AND ESTIMATION**

The model formulation in the current paper takes the form of a latent segmentation duration model. The segmentation is based on the individual's shopping activity regularity, which is unobserved (latent) to the analyst. Each individual is assumed to be either a regular shopper or an erratic shopper. However, since this information is not available to the analyst, the analyst can only assign individuals to the regular and erratic categories probabilistically. In our formulation, the assignment is based on the characteristics of the individual. Within each of the regular and erratic shopper segments, individuals differ in their intershopping duration (*i.e.*, frequency of participation in shopping) based on both observed and unobserved individual characteristics.

The next section discusses the duration model formulation for erratic shoppers, while Section 3.2 presents the corresponding formulation for regular shoppers. Section 3.3 introduces the concept of latent segmentation. Finally, Section 3.4 presents the overall estimation procedure.

#### **3.1 Duration Model for Erratic Shoppers**

An individual is designated an erratic shopper if their likelihood of participation in shopping on any particular day is independent of the time elapsed since the last shopping participation. This implies an exponential distribution for the individual's intershopping duration. In the context of duration modeling, the exponential intershopping duration can be represented in the form of a constant hazard of shopping activity participation, where the hazard on the  $t$ th day since the last shopping participation,  $\lambda(t)$ , is defined as the conditional probability that the individual will

participate in shopping on the  $t$ th day, given that the individual has not participated in shopping before the  $t$ th day. Mathematically, we can write the following:

$$\lambda(t) = \lambda_0 \tag{1}$$

where  $\lambda_0$  represents the constant hazard.

Equation (1) assumes that the intershopping duration distribution is the same across all erratic shoppers (*i.e.*, that the shopping participation frequency is constant across erratic shoppers). However, this is unlikely to be the case in reality, where some erratic shoppers might participate more frequently in shopping than others due to various individual-related characteristics, some of which may be observed by the analyst and some not. To accommodate the effect of observed and unobserved individual characteristics (or covariates), we use a proportional hazard formulation that specifies the effect of covariates to be multiplicative on the baseline constant hazard  $\lambda_0$  (see Bhat, 2000b):

$$\lambda(t) = \lambda_0 \exp(-\beta'x - v), \tag{2}$$

where  $x$  is a vector of covariates specific to each individual,  $\beta$  is a corresponding vector of coefficients to be estimated, and  $v$  is an individual-specific unobserved factor (in the current presentation, we are suppressing the index for individuals for ease in presentation;  $v$  may represent factors such as intrinsic liking or aversion for shopping, and the quality of stores around an individual's residence). The exponential specification for the effect of observed and unobserved covariates in Equation (2) guarantees the positivity of the hazard function without placing constraints on the sign of the elements of the vector  $\beta$ .

The proportional hazard formulation of Equation (2) can be written in the following equivalent form (see Bhat, 2000b):

$$\ln \int_{r=0}^t \lambda_0 dr = \ln(\lambda_0 t) = \beta'x + v + \varepsilon, \quad (3)$$

where  $\varepsilon$  is a erratic term with an extreme value distribution:  $\text{Prob}(\varepsilon < z) = G(z) = 1 - \exp[-\exp(z)]$ . Now let  $k$  be the actual intershopping duration (in days) associated with a particular intershopping spell. Then, using Equation (3), the probability that the intershopping spell duration is  $k$  may be written, conditional on  $v$ , as:

$$\begin{aligned} \text{Prob}[t = k | v] &= \text{Prob}[\ln(\lambda_0(k-1)) < \ln(\lambda_0 t) < \ln(\lambda_0(k))] \\ &= G[\ln(\lambda_0 k) - \beta'x - v] - G[\ln(\lambda_0(k-1)) - \beta'x - v] \end{aligned} \quad (4)$$

Equation (4) is the appropriate probability expression when there are ties within each period  $t$ . We now introduce the index  $q$  for individuals ( $q = 1, 2, \dots, Q$ ) and the index  $i$  for the intershopping duration spell ( $i = 1, 2, \dots, I_q$ ). Then, the likelihood function for an individual  $q$  with  $I_q$  intershopping duration spells can be written, conditional on  $v_q$ , as:

$$L_{q, \text{erratic}} | v_q = \prod_{i=1}^{I_q} \{ [G[\ln(\lambda_0 k_{qi}) - \beta'x_q - v_q] - G[\ln(\lambda_0(k_{qi}-1)) - \beta'x_q - v_q]] \}, \quad (5)$$

where  $x_q$  is the covariate vector for individual  $q$ , and  $k_{qi}$  is the actual intershopping duration of individual  $q$  in the  $i$ th spell. Since individual observations used in the estimation are a sample drawn from the larger population of the residents of Halle and Karlsruhe, it is appropriate to consider the  $v_q$  terms ( $q = 1, 2, \dots, Q$ ) to be erratic variables. We assume the  $v_q$  terms to be normally distributed, with a mean of zero (a normalization) and standard error of  $\sigma^1$ . The

---

<sup>1</sup> For cross-sectional analysis, a convenient distribution for  $\exp(v_q)$  is the gamma distribution, since this leads to a closed form expression when the random component is conditioned out (see Bhat, 1996). However, for panel data, a closed form expression can be obtained only if  $2^{I_q}$  multiplication operations are undertaken for each individual (*i.e.*, the conditional likelihood function of Equation 5 needs to be written as the sum of  $2^{I_q}$  terms). In the current

unconditional likelihood function for individual  $q$  with  $I_q$  intershopping durations, given that the individual is an erratic shopper, can then be written as follows:

$$L_{q,erratic} = \int_{v_q=-\infty}^{\infty} \left[ \prod_{i=1}^{I_q} \{ [G[\ln(\lambda_0 k_{qi}) - \beta' x_q - v_q]] - G[\ln(\lambda_0 (k_{qi} - 1)) - \beta' x_q - v_q] \} \right] dF(v_q | \sigma), \quad (6)$$

where  $F$  represents the cumulative normal distribution function.

### 3.2 Duration Model for Regular Shoppers

An individual is designated as a “regular” shopper if the individual is not an erratic shopper, that is, if the individual’s likelihood of participation in shopping is not independent of the time elapsed since the last shopping participation. We use a general non-parametric hazard to capture the dynamics of intershopping duration. To formulate such a non-parametric hazard model, consider again the proportional hazard form:

$$\eta(t) = \eta_0(t) \exp(-\gamma y - w), \quad (7)$$

where  $\eta(t)$  is the hazard at time  $t$ ,  $\eta_0(t)$  is the baseline hazard at time  $t$ ,  $y$  is a vector of covariates specific to each individual,  $\gamma$  is a corresponding vector of coefficients to be estimated, and  $w$  is an individual-specific unobserved factor. Equation (7) can be rewritten in the following equivalent form:

$$\ln \int_{s=0}^t \eta_0(s) ds = \delta_t = \gamma' y + w + \xi, \quad (8)$$

where  $\xi$  is an erratic term with an extreme value distribution.

---

analysis, this is extremely computation-intensive since some individuals have as many as 36 intershopping durations (this would require  $6.87 \times 10^{10}$  multiplication operations!). The normal distribution assumption used here, on the other hand, is much more computationally feasible.

After introducing the index  $q$  for individuals and the index  $i$  for the intershopping duration spell, and following the same procedure as in the previous section, the unconditional likelihood for individual  $q$  with  $I_q$  durations, given that the individual is a regular shopper, may be written as follows:

$$L_{q,regular} = \int_{w_q=-\infty}^{\infty} \left[ \prod_{i=1}^{I_q} \{ [G[\delta_{k_{qi}} - \gamma' y_q - w_q] - G[\delta_{k_{qi}-1} - \gamma' y_q - w_q]] \} \right] dF(w_q | \theta), \quad (9)$$

where  $\theta$  is the standard error of  $w_q$  and  $k_{qi}$  is the actual intershopping duration of individual  $q$  in the  $i$ th spell.

### 3.3 Latent Segmentation

The previous two sections have derived the likelihood functions corresponding to regular and erratic shoppers. However, the analyst does not have information on whether an individual is a regular shopper or an erratic shopper. So, we endogenously classify individuals as a regular or erratic shopper within a latent segmentation approach (see Bhat, 1997). In such an approach, each individual is probabilistically assigned to the two segments based on individual-related characteristics. Assuming a binary logit structure, the probability that an individual  $q$  is a regular shopper can be written as<sup>2</sup>:

$$P_{q,regular} = \frac{1}{1 + e^{-\alpha' m_q}}, \quad (10)$$

where  $m_q$  is a vector of individual-related characteristics, and  $\alpha$  is a corresponding vector of parameters to be estimated.

---

<sup>2</sup> The logit and probit structures are the two most commonly used discrete choice formulations for binary choice. Both these structures provide almost identical results after appropriate normalizations to account for the different scales of the error terms. In the current paper, we use a binary logit because it has a closed form structure.

The final unconditional likelihood function for individual  $q$  is the weighted average of the likelihood function conditional on the individual being an erratic shopper and the likelihood function conditional on the individual being a regular shopper:

$$L_q = (1 - P_{q,regular}) \cdot L_{q,erratic} + P_{q,regular} \cdot L_{q,regular}. \quad (11)$$

The log-likelihood function to be maximized is  $\sum_q \log L_q$ . Maximization of this function is accomplished using the GAUSS matrix programming language. A Gaussian quadrature approach is used to evaluate the single-dimensional integrals in the likelihood function (see Bhat, 2000b for a discussion of the Gaussian quadrature approach). The parameters to be estimated include the following: (a)  $\lambda_0$ , the vector  $\beta$ , and  $\sigma$  in the duration model for the erratic shopper segment, (b) the scalars  $\delta_t$  ( $t = 1, 2, \dots, T$ ), the vector  $\gamma$ , and  $\theta$  in the duration model for the regular shopper segment, and (c) the vector  $\alpha$  in the latent segmentation model.

## 4. THE DATA

### 4.1 Data Source

The data source for the current study is a 6-week travel survey conducted in Karlsruhe (West Germany) and Halle (East Germany) as part of the MobiDrive study funded by the German Ministry for Research and Education (see Axhausen *et al.*, 2002, for a detailed description of this data source). The main objective of this travel survey data collection was to facilitate a better understanding of the rhythms, routines, and habits of individuals over an extended time period of several weeks. The data collection effort was initiated by contacting a sample of households erratically selected from a phonebook database in each of the two cities. A sub-sample of this larger sample of households was selected for administration of the travel survey, based on eligibility considerations and willingness to participate (only households who did not plan to take

a vacation of more than a week during the survey period and who did not have children under the age of 6 years were deemed eligible).

The collection and analysis of multiday travel data poses several challenging issues. For example, there may be a systematic bias in the types of households that participate in the survey. Further, there is the possibility of (a) households dropping out after a few days, (b) item non-response, and (c) fatigue in reporting during the course of the six-week survey period. As indicated by Axhausen *et al.* (2002), however, these issues do not appear to be significant in the context of the MobiDrive data. Specifically, Axhausen *et al.* indicate that there are no substantial differences in key socio-demographics between respondent and non-respondent households. Households very rarely dropped out from the survey during the six-week period, a finding similar to the experience of the Uppsala study in 1971. Further, item non-response was literally non-existent, in part because respondents were called back over the telephone to clarify errors and ambiguities. Finally, Axhausen *et al.* (2002) and Frascini and Axhausen (2001) have examined fatigue effects, both descriptively and using formal modeling techniques. Their studies do not find any significant evidence of fatigue effects in several dimensions of travel, including the shares of mobile and immobile days within each week, and the number of weekly reported work and non-work journeys and trips.

The final sample from the survey included information on 361 individuals from 162 households. Of these, 44 individuals from 23 households in Karlsruhe participated in a pretest survey, and 317 individuals from 139 households in Karlsruhe and Halle participated in the main survey. The structure and administration procedures were identical in the two surveys. Both the pretest and main surveys were conducted in two waves to capture some amount of seasonal variation. The pretest travel survey was administered between May 31st and July 25th, and the

main survey was administered between September 13th and November 14th. In addition to the six-week continuous travel diary, information on the sociodemographic characteristics of households and their members, car fleet size and composition, and attitudes toward different modes of transport was also collected.

## 4.2 Sample Used and Description

The sample used in the current analysis comprises 3,288 intershopping duration spells of 285 adult individuals (an adult individual is defined as one whose age is equal to or over 16 years; we restricted the empirical analysis to adult individuals on the basis that children are likely to be accompanied by adults for grocery shopping activities, and are not likely to be decision-makers themselves). The number of intershopping duration spells over the course of the survey varies between 1 and 36 across individuals, with an average of 11 spells. The length of the intershopping duration varies between 1 and 33 days. However, there were very few durations larger than 17 days, and so we collapsed all these intershopping durations to 17 days.

Table 1 provides descriptive information on intershopping duration. The column titled “Risk Set” provides information on the number of intershopping episodes which are “at risk” of termination in period  $t$ ; that is, it provides the number of episodes whose length extends at least up to the beginning of period  $t$ . The column labeled “# terminated” indicates the number of episodes that terminate in period  $t$ . The sample hazard associated with each period is computed using the Kaplan-Meier (KM) nonparametric estimator as the number of terminated episodes in period  $t$  divided by the risk set in period  $t$  (see Kiefer, 1988). The sample hazard values are provided in the fourth column of Table 1, and are also plotted in Figure 1. The hazard is quite high in the first four days, perhaps reflecting a tendency of short intershopping durations among

erratic shoppers. In addition the hazard reveals spikes at 7, 10, 12, and 14 days, presumably indicating rhythmic interpurchase durations among regular shoppers. The latent segmentation approach used in the current analysis can separate out the hazard functions of regular and erratic shoppers, as we discuss in the empirical results section of the paper.

### **4.3 Variable Specifications**

The choice of variables for potential inclusion in the model was guided by previous research on shopping trip regularity and frequency, and by intuitive arguments regarding the effect of exogenous variables on shopping activity participation. Three broad sets of variables were considered: individual and spouse characteristics, household characteristics, and location and trip-making characteristics (seasonality and day of week variables were also considered, but did not significantly impact intershopping duration).

Individual and spouse characteristics explored in our specifications included dummy variables for sex, ethnicity, education level, and employment status, and linear and non-linear representations of work hours per week and age. Household characteristics considered in the model included household size, family structure, the number and employment status of household adults, household income, household tenure status (own or rent), household dwelling type (single family unit, duplex, apartment, *etc.*), number of vehicles, and communication-related connections (such as number of telephones, number of private e-mail addresses, and number of fax machines). Location and trip-making characteristics included whether the household is located in Karlsruhe or Halle, the most frequently used mode for shopping activity participation, and the percentage of shopping episodes chained with other activities.

We arrived at the final specification based on a systematic process of eliminating variables found to be insignificant in previous specifications and based on considerations of parsimony in representation. Table 2 provides a list of exogenous variables included in the final specification and their descriptive statistics in the sample.

## **5. EMPIRICAL RESULTS**

### **5.1 Overall Results**

The log-likelihood value at convergence for the latent segmentation model is  $-5612.6$ . The log-likelihood value for the model that considers all individuals to be erratic shoppers is  $-5899.5$  (this model includes the constant hazard corresponding to an exponential intershopping duration distribution, covariates, and the heterogeneity term). The log-likelihood for the corresponding model that considers all individuals to be regular shoppers is  $-5829.0$  (this model includes the parameters corresponding to the non-parametric threshold, covariates, and the heterogeneity term). A likelihood ratio test of the latent segmentation model with the “pure” erratic shopper model clearly indicates that not all shoppers are erratic in the regularity of their shopping participation (the likelihood ratio test statistic is 573.8, which is larger than the chi-squared statistic with 32 degrees of freedom at any reasonable level of significance). Similarly, a likelihood ratio test of the latent segmentation model with the “pure” regular shopper model indicates that not all shoppers are routine in the regularity of their shopping participation (the likelihood ratio test statistic is 432.8, which is larger than the chi-squared statistic with 21 degrees of freedom at any level of significance). Clearly, these test statistics demonstrate the need to recognize differences in shopping activity participation regularity across individuals.

The difference in intershopping duration between regular and erratic shoppers may be observed by computing the mean intershopping duration in each segment. The mean duration in the erratic segment may be computed as:

$$\bar{t}_{erratic} = \frac{\sum_q \sum_{i=1}^{I_q} P_{q,erratic} \cdot \left\{ \sum_l [\text{Prob}(t_{qi} = l) | \text{erratic segment}] \cdot l \right\}}{\sum_q I_q \cdot P_{q,erratic}}, \quad (12)$$

where  $\text{Prob}(t_{qi} = l) | \text{erratic segment}$  is the probability that the  $i$ th duration spell of individual  $q$  is of length  $l$ , given that the individual is in the erratic segment. This probability is given by:

$$\text{Prob}(t_{qi} = l) | \text{erratic segment} = \frac{\int_{-\infty}^{+\infty} \{G[\ln(\lambda_0 l) - \beta' x_{qi} - v_q] - G[\ln(\lambda_0(l-1)) - \beta' x_{qi} - v_q]\} dF(v_q | \sigma)}{\int_{-\infty}^{+\infty} G[\ln(\lambda_0 l) - \beta' x_{qi} - v_q] dF(v_q | \sigma)} \quad (13)$$

The mean intershopping duration in the regular segment may be similarly computed. These values are estimated to be 1.8 days for the erratic segment and 3.6 days for the regular segment. Thus, the intershopping duration for erratic shoppers tends to be about half that of regular shoppers (*i.e.*, erratic shoppers participate, on average, about twice as frequently in shopping as regular shoppers).

The next section discusses the baseline hazard estimates in each of the erratic and regular segments. Section 5.3 interprets the covariate effects in the erratic and regular shopper segments. Section 5.4 presents the unobserved heterogeneity estimates. Finally, Section 5.5 presents the results of the segmentation model for erratic versus regular shoppers. It is important to note that the parameters discussed in the various sections are all estimated simultaneously. They are discussed separately for ease in presentation.

## 5.2 Baseline Hazard

The baseline hazard function for the erratic shopper segment is constant across intershopping periods with a value of 0.912. This hazard parameter is highly statistically significant with a  $t$ -value of 8.70. The constant hazard reflects the “memoryless” property of erratic shoppers; that is, the propensity to participate in shopping at time  $t$  (given that the individual does not go shopping until time  $t$ ) is independent of time  $t$ .

The baseline hazard function for regular shoppers is shown in Figure 2. The baseline hazard indicates a primary spike at 7 days and other secondary spikes at 10, 12, and 14 days, similar to the sample hazard in Figure 1. The baseline hazard also exhibits non-monotonic behavior like the sample hazard. However, the baseline hazard for regular shoppers is different from the sample hazard in important ways (note that the sample hazard does not distinguish between erratic and regular shoppers). First, the baseline hazard is increasing between 1 and 4 days, while the sample hazard is decreasing during the same time window. That is, the propensity of shopping activity participation for regular shoppers increases as the time elapsed since the last participation increases from 1 to 4 days; however, the probability of shopping activity participation for all shoppers decreases during the same period. Second, the hazard function is lower for the baseline hazard in the first 3 days compared to the sample hazard. This latter finding is a consequence of the smaller intershopping duration of erratic shoppers compared to regular shoppers, resulting in a higher hazard in the early periods when all shoppers are considered together. Of course, the difference in the hazard functions could also be attributed to covariate effects (note that the sample hazard does not include covariates). Third, the spike at 7 days is much more pronounced for regular shoppers than for all shoppers. This indicates a strong weekly rhythm of shopping activity participation for regular shoppers. Fourth, the spikes

at 10, 12, and 14 days are also more pronounced in the baseline hazard of regular shoppers than in the sample hazard.

To summarize, two general conclusions may be drawn from the above results. First, the baseline hazard for erratic shoppers is higher than the baseline hazard for regular shoppers for all time periods. Thus, in general, erratic shoppers have shorter intershopping durations than regular shoppers. This is consistent with the finding in the previous section. Second, there is a clear rhythmic pattern of shopping participation for regular shoppers, with a large spike at 7 days, and secondary spikes at 10, 12, and 14 days. Thus, the hazard function for regular shoppers is not smooth and is not monotonic. Parametric baseline hazard functions used commonly in transportation research cannot capture such a hazard profile; a non-parametric baseline approach, which is also able to handle the discrete nature of time periods and ties within each period, is the appropriate approach.

### 5.3 Covariate Effects

In this section, we discuss the effects of covariates on the duration hazard. It should be observed from Equations (2) and (7) that a positive coefficient on a covariate implies that the covariate lowers the hazard rate, or equivalently, the covariate increases the intershopping duration. The magnitude of the effect of covariates within each of the erratic and regular segments may be assessed by computing the percentage change in the hazard due to a one unit change in the covariate. From the functional form in Equation (2), this can be written as  $\{\exp(\beta_k) - 1\} \times 100$  for a one-unit change in  $x_k$ , the  $k$ th covariate in the erratic shopper segment. Similarly, the percentage change in the hazard due to a one unit change in the  $k$ th covariate in the regular shopper segment may be written using Equation (7) as  $\{\exp(\lambda_k) - 1\} \times 100$ .

Table 3 shows the estimated covariate effects for the final model specification. These effects are discussed by variable category in the subsequent three paragraphs.

The effect of individual and spouse characteristics indicates that individuals who work longer have a lower hazard (*i.e.*, a higher intershopping duration) than individuals who work shorter durations. This applies to both the erratic and regular segment groups, and might be a manifestation of tighter time constraints for individuals who work long hours. Older individuals have a lower hazard compared to younger individuals in the erratic shopper segment, perhaps due to lower mobility levels (nonlinear effects of age were also tested, but did not improve data fit substantially and were also difficult to interpret). Finally, within the category of individual and spouse characteristics, spousal employment leads to a higher hazard for both segments, possibly due to higher responsibility for household maintenance activities if an individual's spouse is employed.

The effects of household characteristics show that erratic shoppers in high income households have a low hazard (high intershopping duration). The coefficient on the "house" variable indicates that regular shoppers living in a house have a lower hazard (higher intershopping duration) than regular shoppers living in multi-family dwelling units, presumably due to the ability to stock-up for longer durations because of larger storage space (or smaller inventory costs).

The effects of location and trip-making characteristics indicate the following: (a) Erratic shoppers in Karlsruhe have a higher hazard (lower intershopping duration) compared to erratic shoppers in Halle, (b) Shoppers who use a car as the primary mode to participate in shopping have a lower hazard than those who use other modes (this may reflect the ability to carry large amounts of groceries if a car is used, resulting in less need to shop frequently), and (c)

Individuals who chain shopping participations with other activity stops are more likely to shop frequently (perhaps due to relative ease of participating in shopping if the activity is chained with other activities).

#### 5.4 Unobserved Heterogeneity

Unobserved heterogeneity across individuals is included in the model using a normal mixing distribution. The unobserved heterogeneity term acts multiplicatively on the hazard function. The variance of this term is estimated to be 0.1226 with a  $t$ -statistic of 3.90 for the erratic segment and 0.0330 with a  $t$ -statistic of 1.76 for the regular segment. Thus, the results indicate the presence of factors other than those included in the model that impact intershopping durations. Such factors may include inherent individual preferences, household responsibilities/interactions, and other miscellaneous unobserved determinants. It is well established now that failure to control for such unobserved heterogeneity can produce severe bias in the nature of duration dependence and the estimates of covariate effects (Heckman and Singer, 1984; Lancaster, 1985).

The magnitude of the variance of the unobserved heterogeneity term provides important information regarding the fraction of the variation in the intershopping hazard rate explained by the covariates and by unobserved factors (see Jain and Vilcassim, 1991). This is most easily observed by taking the logarithm of both sides of Equation (2) and Equation (7), yielding the following equations:

$$\ln \lambda(t) = \ln \lambda_0 - \beta'x - v, \text{ and } \ln \eta(t) = \ln \eta_0(t) - \gamma'y - w. \quad (14)$$

Since the baseline hazards,  $\lambda_0$  and  $\eta_0(t)$  are the same across individuals in the erratic shopper and regular shopper segments, respectively, the variance across individuals of the (log)

intershopping time hazard is simply the variance across individuals due to observed heterogeneity and the variance across individuals due to unobserved factors (assuming that the covariates and unobserved heterogeneity are uncorrelated):

$$\text{Var}[\ln \lambda(t)] = \text{Var}[\beta'x] + \text{Var}[v], \text{ and } \text{Var}[\ln \eta(t)] = \text{Var}(\gamma'y) + \text{Var}(w). \quad (15)$$

The fraction of variation in the departure time hazard explained by unobserved heterogeneity may then be computed as  $\text{Var } v / (\text{Var}[\beta'x] + \text{Var}[v])$  for the erratic shopper segment and as  $\text{Var } w / (\text{Var}[\gamma'y] + \text{Var}[w])$  for the regular shopper segment. This fraction is about 20% for the erratic segment and only about 2% for the regular segment. Clearly, the results indicate that it is substantially more difficult to explain the shopping rates of erratic shoppers than it is to explain the rates of regular shoppers, based on observed sociodemographic and other attributes.

## 5.5 Segmentation Model Results

Table 4 provides the segmentation model results for probabilistically assigning individuals into the erratic and regular shopper segments. The parameter estimates indicate the propensity to belong to the regular shopper segment relative to the erratic shopper segment. The results corresponding to individual characteristics in Table 4 indicate that men are more likely to be regular shoppers than women, while individuals with a high education level are more likely to be erratic shoppers. The latter result might be a reflection of tight schedule constraints among highly educated individuals, resulting in participation in shopping on a “as time permits” basis rather than a regularly scheduled pattern.

The parameters on household characteristics indicate that individuals in nuclear family and couple family households are more likely to be erratic shoppers than individuals in other household types (the single person household dominates the other household types). This result

may be attributed to the higher likelihood of stocking out of grocery items in the presence of small children and other adults in the household, due to the diversity of food consumption needs and higher variation in consumption levels. Consequently, such households are more likely to undertake “fill-up” shopping trips, which tend to be erratic. The final two parameters under household characteristics indicate that individuals living in a house, and in households with several vehicles, tend to be regular shoppers. Individuals living in a house have more storage space, and thus can stockpile groceries depending on their desired shopping participation rhythm. On the other hand, individuals in apartments and other multifamily household units have limited storage space, resulting in the need for erratically spaced “fill-up” shopping trips due to the higher likelihood of stocking out. The finding that individuals in households with several vehicles tend to be regular shoppers may reflect the effect of vehicle availability on shopping regularity. Fewer vehicles in the household would imply more competition for vehicles and less control over times when a vehicle is available for personal use, translating to erratic shopping trips.

Finally, the results in Table 4 indicate that residents of Karlsruhe are more erratic in their shopping patterns than residents of Halle. This may be a manifestation of the better economic conditions in West Germany compared to East Germany, because of which residents of Karlsruhe make “spur-of-the-moment” shopping trips, while residents of Halle systematically “ration” their shopping activity participation

An intuitive way to characterize the attributes of each of the erratic and regular shopper segments is to compute the mean of the exogenous variables in Table 4 in each segment as follows:

$$\bar{m}_{erratic} = \frac{\sum_q P_{q, erratic} m_q}{\sum_q P_{q, erratic}} \text{ and } \bar{m}_{regular} = \frac{\sum_q P_{q, regular} m_q}{\sum_q P_{q, regular}}. \quad (16)$$

The size of each of the segments (in terms of share) can be similarly computed as:

$$R_{erratic} = \frac{\sum_q P_{q, erratic}}{Q} \text{ and } R_{regular} = \frac{\sum_q P_{q, regular}}{Q}. \quad (17)$$

The results are presented in Table 5. The means of the exogenous variables in the two segments, and in the entire sample of shoppers, support our previous observations regarding segment characteristics. The last row in Table 5 indicates that about 28% of the individuals are erratic shoppers, while 72% are regular shoppers; that is, the split between erratic and regular shoppers is about 1:2.6.

## 6. CONCLUSIONS

This paper examines the rhythms in the shopping activity participation of individuals over a multiweek period using a continuous six-week travel survey collected in the cities of Halle and Karlsruhe in Germany in the fall of 1999. The rhythms in shopping activity participation are examined by modeling the duration between successive shopping activity participations of individuals. The methodology uses a hazard-based duration model structure that allows different individuals to have different rhythms in behavior and that is able to predict shopping activity participation behavior (both frequency and distribution of the activity participations) over any period of time (such as a day, a week, or a month). An important characteristic of the proposed formulation is that it endogenously classifies individuals into erratic and regular shoppers based on individual characteristics, and models intershopping duration in each group as a function of both observed and unobserved individual characteristics. To our knowledge, this is the first

application of a latent segmentation scheme for duration modeling that accommodates the effect of explanatory variables on segment membership as well as on intershopping duration.

The paper provides several important insights into the determinants of the regularity and frequency of shopping activity behavior. First, the results indicate that shoppers differ both in their regularity as well as frequency of shopping activity participation. Erratic shoppers tend to have a shorter intershopping duration (higher shopping frequency) than regular shoppers. Second, the intershopping duration dynamics of regular shoppers shows a very clear weekly trend in shopping participation; in addition, intershopping durations of 12 and 14 days are also quite prevalent. A related result is that the baseline hazard function for regular shoppers is not smooth and is not monotonic. Thus, it is important to use a nonparametric hazard approach rather than parametric hazard shapes for intershopping duration. Third, the empirical results indicate the strong influence of individual and spousal employment-related attributes, travel mode used for shopping, and trip-chaining behavior on shopping frequency. Fourth, the results show that the hazard model performs quite well, with observed factors explaining about 80% (98%) of the overall variation in the departure time hazard for erratic (regular) shoppers. Clearly, it is easier to explain the hazard variation among regular shoppers than among erratic shoppers. Fifth, erratic and regular shoppers are systematically different in their characteristics. Specifically, women, highly educated individuals, individuals in couple and nuclear family households, and individuals living in a house and whose households have several vehicles are likely to be erratic shoppers.

An examination of the regularity and frequency of activity participation over an extended period of multiple days is important for accurate transportation policy analysis, as emphasized earlier in the paper. This is particularly the case at a time when population demographics,

household structures, and employment characteristics are changing rapidly in the population. For example, the percentage of the German population above 60 years is projected to climb from 23.2% in 2000 to 38.1% in 2050. The shift in age distribution, combined with the longer lifespan of women (relative to men), is projected to affect the gender distribution in the German population (see Population Division, United Nations, 2000). In addition, the Federal Statistical Office of Germany indicates an increasing trend toward employment, primarily because of higher labor force participation of women (the labor force participation of women has increased from 63.8% in 1999 to 65% in 2001, with the trend projected to continue; see Federal Statistical Office Germany, 2002). Women are also increasing their educational qualifications and their work hours, according to the Federal Statistical Office. These various shifts in age, sex distribution, employment, educational qualifications, and work hours will result in shifts in the regularity and frequency of shopping activity participation, which can be predicted by the model developed in the paper.

The paper contributes toward such a multiday examination of activity behavior in the context of shopping participation of individuals. However, the paper is not without its limitations. First, the current paper focuses only on intershopping duration for maintenance-related shopping, without considering the potential joint nature of such decisions with participation in other types of activities (such as recreation and non-maintenance shopping). Second, the effect of interaction among household members on individual activity behavior is represented in the form of simple measures such as marital status, spouse's employment characteristics, and household structure. However, to explicitly accommodate the interaction among household members, it would be more appropriate to view individual activity behavior within the broader context of household activity behavior, and to model the activity behavior of

all individuals in a household jointly. Third, it may be more appropriate to consider certain “exogenous” variables in the current analysis as being co-determined with intershopping duration. For example, the need to shop frequently may drive the decision to use the car mode and/or to chain shopping activities with other activities. Thus, it would be useful to model travel mode choice, activity chaining behavior, and intershopping time jointly. Further extension to model other dimensions of multiday activity participation behavior, such as time of day of travel, duration of activity stops, travel mode used, and location of activity stops, will also be useful, though it also promises to be challenging.

#### **ACKNOWLEDGEMENT**

The authors would like to thank two anonymous reviewers for their constructive suggestions on an earlier version of the paper. The authors are also grateful to Lisa Weyant for her help in typesetting and formatting this document.

## REFERENCES

- Arentze, T.A., and H.J.P. Timmermans (2002). Albatross – A Learning-Based Transportation Oriented Simulation System. Forthcoming in *Transportation Research*.
- Axhausen, K.W., Zimmerman, Z., Schönfelder, S., Rindsfuser, G. and T. Haupt (2002). Observing the Rhythms of Daily Life: A Six Week Travel Diary. *Transportation*, Vol. 29, No. 2, pp. 95-124.
- Bhat, C.R. (1996). A Hazard-Based Duration Model of Shopping Activity with Nonparametric Baseline Specification and Nonparametric Control for Unobserved Heterogeneity. *Transportation Research*, Vol. 30B, No. 3, pp. 189-207.
- Bhat, C.R. (1997). An Endogenous Segmentation Mode Choice Model with an Application to Intercity Travel. *Transportation Science*, Vol. 31, No. 1, pp. 34-48.
- Bhat, C.R. (1999). An Analysis of Evening Commute Stop-Making Behavior Using Repeated Choice Observations from a Multiday Survey. *Transportation Research*, Vol. 33B, No. 7, pp. 495-510.
- Bhat, C.R. (2000a). Incorporating Observed and Unobserved Heterogeneity in Urban Work Travel Mode Choice Modeling. *Transportation Science*, Vol. 34, No. 2, pp. 228-238.
- Bhat, C.R. (2000b). Duration Modeling. In *Handbook of Transport Modelling*, (D.A. Hensher and K.J. Button, eds), Elsevier Science, pp. 91-111.
- Bhat, C.R. and F.S. Koppelman (1999). Activity-Based Modeling of Travel Demand. In *The Handbook of Transportation Science*, (Randolph Hall, ed.), Kluwer Academic Publishers, Norwell, Massachusetts, pp. 35-61.
- Bhat, C.R. and S.K. Singh (2000). A Comprehensive Daily Activity-Travel Generation Model System for Workers. *Transportation Research*, Vol. 34A, No. 1, pp. 1-22.
- Bowman, J.L., and M.E. Ben-Akiva (2000). Activity-Based Disaggregate Travel Demand Model System with Activity Schedules. *Transportation Research*, Vol. 35A, pp. 1-28.
- Federal Statistical Office Germany (2002). Employment, [www.destatis.de/themen/e/thm\\_employment.htm](http://www.destatis.de/themen/e/thm_employment.htm)
- Fraschini, E. and K.W. Axhausen (2001). Day on Day Dependencies in Travel: First Results Using ARIMA Modeling. *Arbeitsberichte Verkehrs- und Raumplanung* 63, Institut für Verkehrsplanung, Transporttechnik, Strassen- und Eisenbahnbau, ETH Zürich, Zürich.

- Han, A. and J.A. Hausman (1990). Flexible Parametric Estimation of Duration and Competing Risk Models. *Journal of Applied Econometrics*, Vol. 5, pp. 1-28.
- Hanson, S. and J.O. Huff (1986). Classification Issues in the Analysis of Complex Travel Behavior. *Transportation*, Vol. 13, pp. 271-293.
- Hanson, S. and J.O. Huff (1988a). Systematic Variability in Repetitious Travel. *Transportation*, Vol. 15, pp. 111-135.
- Hanson, S. and J.O. Huff (1988b). Repetition and Day-to-Day Variability in Individual Travel Patterns: Implications for Classification. In *Behavioral Modelling in Geography and Planning*, (R. Golledge and H. Timmermans, eds), Croom Helm, New York.
- Hatcher, G. and H.S. Mahmassani (1992). Daily Variability of Route and Trip Scheduling Decisions for the Evening Commute. *Transportation Research Record 1357*, TRB, National Research Council, Washington, D.C., pp. 72-81.
- Heckman, J. and B. Singer (1984). A Method for Minimizing the Distributional Assumptions in Econometric Models for Duration Data. *Econometrics*, Vol. 52, pp. 271-320.
- Hensher, D.A. and F.L. Mannering (1994). Hazard-Based Duration Models and Their Application to Transport Analysis. *Transport Reviews*, Vol. 14, No. 1, pp. 63-82.
- Huff, J.O. and S. Hanson (1986). Repetition and Variability in Urban Travel. *Geographical Analysis*, Vol. 18, No. 2, pp. 97-113.
- Huff, J.O. and S. Hanson (1990). Measurement of Habitual Behavior: Examining Systematic Variability in Repetitive Travel. In *Developments in Dynamic and Activity-Based Approaches to Travel Analysis*, (P. Jones, ed.), Gower Publishing Co., Aldershot, England, pp. 229-249.
- Jain, D.C. and N.J. Vilcassim (1991). Investigating Household Purchase Timing Decisions: A Conditional Hazard Function Approach. *Marketing Science*, Vol. 10, pp. 1-23.
- Jones, P. and M. Clarke (1988). The Significance and Measurement of Variability in Travel Behaviour. *Transportation*, Vol. 15, pp. 65-87.
- Kalbfleisch, J.D. and R.L. Prentice (1980). *The Statistical Analysis of Failure Time Data*. John Wiley and Sons, New York, p. 101
- Kiefer, N.M. (1988). Econometric Duration Data and Hazard Functions. *Journal of Econometric Literature*, Vol. 27, pp. 646-679.

- Kim, B. and K. Park (1997). Studying Patterns of Consumer's Grocery Shopping Trip. *Journal of Retailing*, Vol. 73, No. 4, pp. 501-517.
- Kitamura, R., and S. Fujii (1998). Two Computational Process Models of Activity Travel Behavior. In Garling, T., Laitila, T., and K. Westin (eds) *Theoretical Foundations of Travel Choice Modeling*, Elsevier Science, Oxford, England.
- Koppelman, F.S. and E.I. Pas (1984). Estimation of Disaggregate Regression Models of Person Trip Generation With Multiday Data. In Volmuller, J. and R. Hamerslag (eds) *Proceedings of the Ninth International Symposium on Transportation and Traffic Theory*, VNU Science Press, Utrecht, The Netherlands, pp. 513-529.
- Lancaster, T. (1985). Generalized Residuals and Heterogeneous Duration Models with Applications to the Weibull Model. *Journal of Econometrics*, Vol. 28, pp. 155-169.
- Mahmassani, H.S. (1997). Dynamics of Commuter Behavior: Recent Research and Continuing Challenges. In *Understanding Travel Behavior in an Era of Change*, 1<sup>st</sup> edition, Pergamon, Elsevier Science, Inc., Tarrytown, NY, pp. 279-313.
- Meyer, B.D. (1990). Unemployment Insurance and Unemployment Spells. *Econometrica*, Vol. 58, No. 4, pp. 757-782.
- Misra, R. and C.R. Bhat (2000). Activity Travel Patterns of Non-Workers in the San Francisco Bay Area: Exploratory Analysis. In *Transportation Research Record 1718*, TRB, National Research Council, Washington, D.C., pp. 43-51.
- Muthyalagari, G.R., Parashar, A. and R.M. Pendyala (2001). Measuring Day-to-Day Variability in Travel Characteristics Using GPS Data. Presented at the 80<sup>th</sup> Annual Meeting of the Transportation Research Board, Washington D.C., January.
- Pas, E.I. (1983). A Flexible and Integrated Methodology for Analytical Classification of Daily Travel-Activity Behavior. *Transportation Science*, Vol. 17, pp. 405-429.
- Pas, E.I. (1988). Weekly Travel-Activity Behavior. *Transportation*, Vol. 15, pp. 89-109.
- 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, pp. 3-20.
- Pas, E.I. and S. Sundar (1995). Intra-personal Variability in Daily Urban Travel Behavior: Some Additional Evidence. *Transportation*, Vol. 22, pp. 135-150.
- Pendyala, R.M., and K.G. Goulias (2002). Time Use and Activity Perspectives in Travel Behavior Research. *Transportation*, Vol. 29, No. 1, pp. 1-4.

- Pendyala, R.M., Yamamoto, T., and R. Kitamura (2002). On the Formulation of Time-Space Prisms to Model Constraints on Personal Activity-travel Engagement. *Transportation*, Vol. 29, No. 1, pp. 73-94.
- Population Division, United Nations (2000). World Population Ageing 1950-2050, Germany, 244-245,  
[www.un.org/esa/population/publications/worldageing19502050/pdf/098germa.pdf](http://www.un.org/esa/population/publications/worldageing19502050/pdf/098germa.pdf)
- Schlich (2001). Analyzing Intrapersonal Variability of Travel Behavior Using the Sequence Alignment Method. Paper presented at the European Transport Conference, Cambridge, September.
- Schönfelder, S. and K.W. Axhausen (2000). Modelling the Rhythms of Travel Using Survival Analysis. Paper submitted to the 80<sup>th</sup> Annual Meeting of the Transportation Research Board, Washington, D.C., January 2001.
- Wen, C. and F.S. Koppelman (1999). Integrated Model System of Stop Generation and Tour Formation for Analysis of Activity and Travel Patterns. In *Transportation Research Record 1676*, TRB, National Research Council, Washington, D.C., pp. 136-144.

**LIST OF FIGURES**

Figure 1. Kaplan-Meier Sample Hazard Function

Figure 2. Baseline Hazard Function for Regular Shoppers

**LIST OF TABLES**

Table 1. Inter-Shopping Duration and Sample Hazard

Table 2. Variable Definitions and Sample Statistics

Table 3. Intershopping Duration Model Results

Table 4. Segmentation Model Results for Erratic Versus Regular Shoppers (Base is the erratic shopper category)

Table 5. Mean Characteristics of Erratic and Regular Shopper Segments

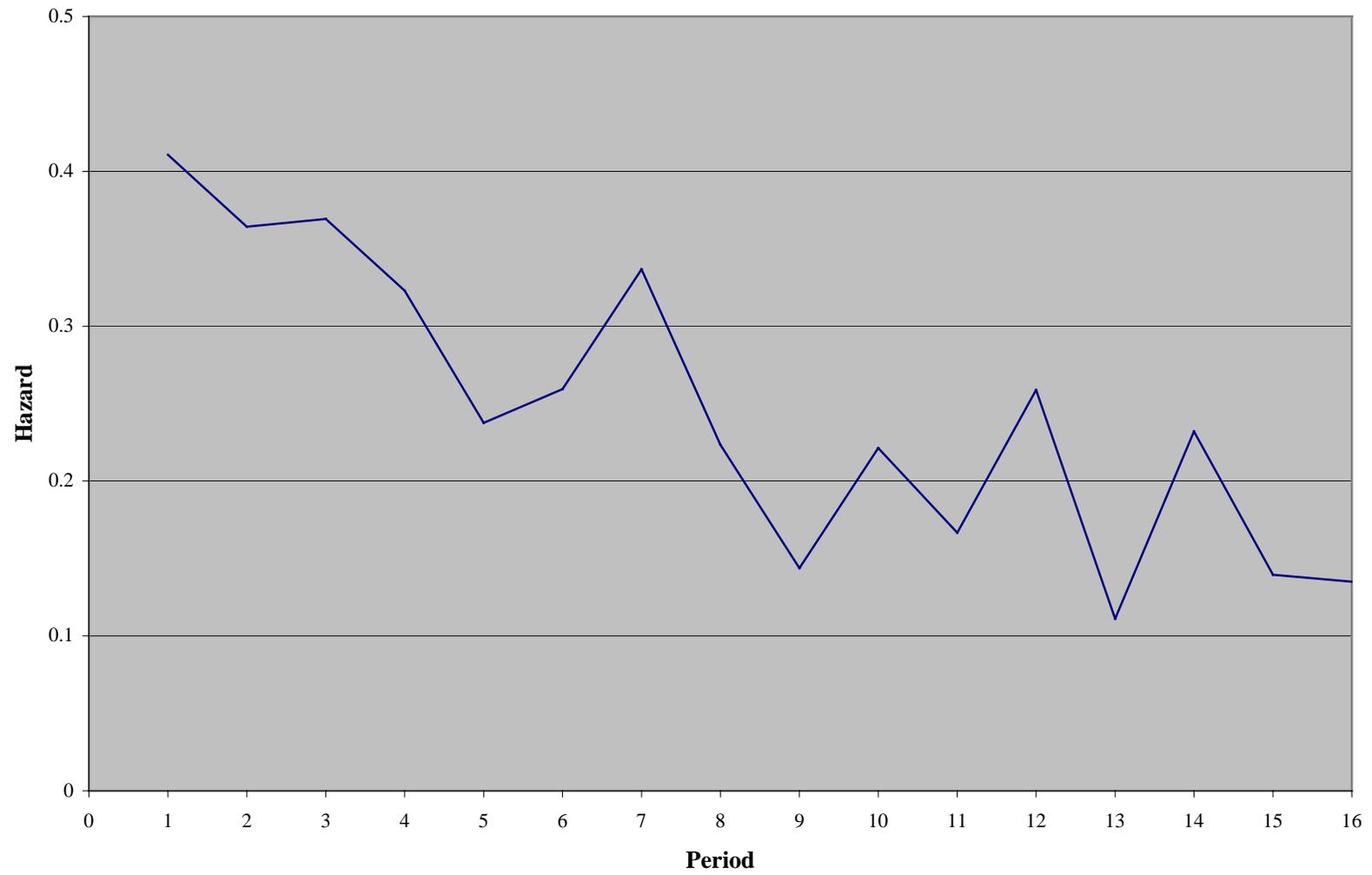


Figure 1. Kaplan-Meier Sample Hazard Function

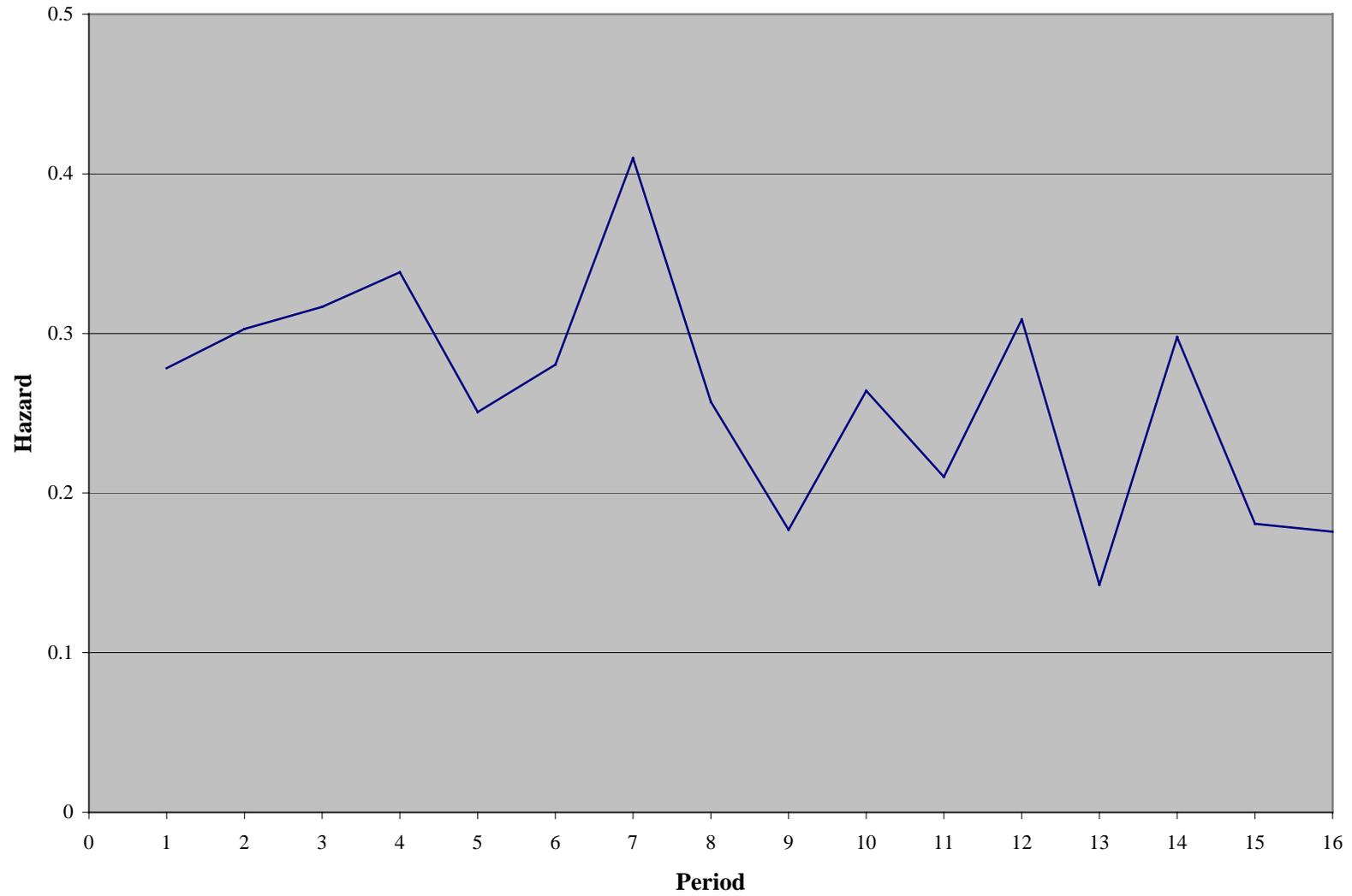


Figure 2. Baseline Hazard Function for Routine Shoppers

**Table 1. Inter-Shopping Duration and Sample Hazard**

| <b>Intershoppping<br/>duration <math>t</math> (in days)</b> | <b>Risk Set</b> | <b># Terminated</b> | <b>Hazard</b> | <b>t-stat</b> |
|---|-----------------|---------------------|---------------|---------------|
| 1   | 3288            | 1350                | 0.4106        | 46.9353       |
| 2   | 1938            | 706                 | 0.3643        | 32.3942       |
| 3   | 1232            | 455                 | 0.3693        | 26.8733       |
| 4   | 777             | 251                 | 0.3230        | 19.2594       |
| 5   | 526             | 125                 | 0.2376        | 12.8052       |
| 6   | 401             | 104                 | 0.2594        | 11.8506       |
| 7   | 297             | 100                 | 0.3367        | 12.2793       |
| 8   | 197             | 44                  | 0.2234        | 7.5268        |
| 9   | 153             | 22                  | 0.1438        | 5.0685        |
| 10  | 131             | 29                  | 0.2214        | 6.1027        |
| 11  | 102             | 17                  | 0.1667        | 4.5162        |
| 12  | 85              | 22                  | 0.2588        | 5.4481        |
| 13  | 63              | 7                   | 0.1111        | 2.8055        |
| 14  | 56              | 13                  | 0.2321        | 4.1143        |
| 15  | 43              | 6                   | 0.1395        | 2.6400        |
| 16  | 37              | 5                   | 0.1351        | 2.4037        |
| 17  | 32              | 32                  | 1.0000        | 1.0000        |

**Table 2. Variable Definitions and Sample Statistics**

| <b>Variable</b>                              | <b>Definition</b>  | <b>Mean</b> | <b>Std. Dev.</b> |
|--|--|-------------|------------------|
| <b>Individual and spouse characteristics</b> |  |             |                  |
| Male   | 1 if individual is male, 0 otherwise   | 0.49        | 0.50             |
| High education level                         | 1 if individual has technical college or university education, 0 otherwise         | 0.31        | 0.46             |
| Spouse employment                            | 1 if spouse is employed, 0 if spouse is not employed or if individual is unmarried | 0.45        | 0.50             |
| Number of work hours                         | Number of work hours per week  | 22.45       | 20.84            |
| Age  | Age of individual (in years)   | 45.91       | 15.80            |
| <b>Household Characteristics</b>             |  |             |                  |
| Nuclear family                               | 1 if family includes parents and 1 or more children, 0 otherwise                   | 0.40        | 0.49             |
| Couple household                             | 1 if household comprises two adults, one male and one female, 0 otherwise          | 0.34        | 0.47             |
| Household income (000s)                      | Monthly household income (in 1000s of Deutsche Marks)                              | 4.12        | 2.00             |
| House  | 1 if dwelling type is single family unit, duplex or a terraced house, 0 otherwise  | 0.34        | 0.72             |
| Number of vehicles                           | Number of vehicles (of all types) in household                                     | 3.41        | 1.92             |
| <b>Location/Trip-Making Characteristics</b>  |  |             |                  |
| Karlsruhe                                    | 1 if household lives in Karlsruhe, 0 otherwise                                     | 0.56        | 0.50             |
| Car is primary mode                          | 1 if car is the most frequently used mode for shopping trips, 0 otherwise          | 0.48        | 0.50             |
| Percentage of shopping episodes chained      | Percentage of shopping episodes chained with other activities                      | 0.43        | 0.31             |

**Table 3. Intershopping Duration Model Results**

| Variable  | Erratic Shopper Segment |                | Regular Shopper Segment |                |
|---|-------------------------|----------------|-------------------------|----------------|
|   | Parameter               | <i>t</i> -stat | Parameter               | <i>t</i> -stat |
| <b>Individual and Spouse Characteristics</b>    |                         |                |                         |                |
| Number of work hours (in 10s)                   | 0.103                   | 4.84           | 0.098                   | 6.88           |
| Age (in 100s)                                   | 0.237                   | 3.20           | --                      | --             |
| Spousal employment                              | -0.104                  | -1.23          | -0.299                  | -4.99          |
| <b>Household Characteristics</b>                |                         |                |                         |                |
| Income (in 1000s)                               | 0.015                   | 1.56           | --                      | --             |
| House   | --                      | --             | 0.047                   | 1.26           |
| <b>Location and Trip-Making Characteristics</b> |                         |                |                         |                |
| Karlsruhe                                       | -0.186                  | -2.06          | --                      | --             |
| Car is primary mode                             | 0.180                   | 1.95           | 0.274                   | 4.63           |
| Percentage of shopping episodes chained         | -0.318                  | -2.36          | -0.142                  | -1.57          |

**Table 4. Segmentation Model Results for Erratic Versus Regular Shoppers**  
(Base is the erratic shopper category)

| Variable                          | Parameter | <i>t</i> -stat |
|-----------------------------------|-----------|----------------|
| <b>Individual Characteristics</b> |           |                |
| Male                              | 0.892     | 2.55           |
| High education level              | -1.294    | -3.18          |
| <b>Household Characteristics</b>  |           |                |
| Nuclear family                    | -2.480    | -3.74          |
| Couple family                     | -1.799    | -3.74          |
| House                             | 0.470     | 1.74           |
| Number of vehicles                | 0.388     | 2.97           |
| <b>Location Characteristics</b>   |           |                |
| Karlsruhe                         | -0.864    | -2.09          |
| <b>Constant</b>                   | 1.691     | 3.42           |

**Table 5. Mean Characteristics of Erratic and Regular Shopper Segments**

| Variable                          | Erratic Shopper Segment | Regular Shopper Segment | All Shoppers |
|-----------------------------------|-------------------------|-------------------------|--------------|
| <b>Individual Characteristics</b> |                         |                         |              |
| Male                              | 0.408                   | 0.524                   | 0.491        |
| High education level              | 0.405                   | 0.271                   | 0.309        |
| <b>Household Characteristics</b>  |                         |                         |              |
| Nuclear family                    | 0.411                   | 0.391                   | 0.396        |
| Couple family                     | 0.450                   | 0.293                   | 0.337        |
| House                             | 0.231                   | 0.378                   | 0.337        |
| Number of vehicles                | 3.145                   | 3.519                   | 3.414        |
| <b>Location Characteristics</b>   |                         |                         |              |
| Karlsruhe                         | 0.544                   | 0.568                   | 0.561        |
| <b>Segment size</b>               | 0.281                   | 0.719                   | 1.000        |