Analysing the rhythms of travel using survival analysis

Paper submitted for presentation at the Transportation Research Board (TRB) 2001 Annual Meeting, Washington 2001 (Call for papers, A1C04(1))

S Schönfelder and KW Axhausen
Abstract

Temporal aspects of individual travel behaviour have been the foci of analysis in the activity-based approach since its emergence 30 years ago. Among the other aspects, rhythmic patterns of travel behaviour have been neglected, though – mainly due to missing or insufficient data capturing periodicity on a weekly or monthly basis. The German research project Mobidrive, funded by the Federal Ministry of Education and Research, implemented a six-week continuous travel diary which now yields a unique data base for detecting and analysing rhythms of time-use and mobility. This paper reports first results describing and explaining those rhythmic patterns of travel behaviour by applying the techniques and models of survival analysis.

Keywords

Rhythms, travel behaviour, periodicity, survival analysis, hazard models, model estimation
1 INTRODUCTION: RHYTHMIC PATTERNS OF TRAVEL BEHAVIOUR

Temporal aspects of individual travel behaviour such as routines, variability or dynamics have been the foci of analysis in the activity-based approach since its emergence thirty years ago (see Kitamura, 1988; Mahmassani, 1988; Pas and Harvey, 1997). A deeper insight into the temporal structures of mobility is believed to further the development of advanced disaggregated transport models and the conceptualisation of transport related policies (Jones, 1981).

The following paper deals with the periodicity of travel behaviour, an aspect of mobility which has been widely neglected. Understanding the rhythmic structure of space-time behaviour is an essential decision background for private and public sector actors in the transport field. Rhythmic patterns structure behavioural elements of travel, such as the complete daily trip or activity chain with matching attributes, activity sequences or single main activities. All of them can be observed in periodic intervals. In this paper, rhythms are considered which extend over prolonged periods of time, such as weeks or months. These rhythms exceed the usual reporting periods of typical travel surveys. The elements to be observed mostly occur on a periodic and predictable basis and can be explained historically, i.e. they are the result of the co-ordination between the human physiology, the dynamic travel environment and the social networks of the travellers (Shapcott and Steadman, 1978).

1.1 Mobidrive – Dynamic and Routines of Travel Behaviour

The analysis of rhythmic patterns of travel behaviour has been so far restricted by the absence of surveys with reporting periods substantially longer than one week. The reporting periods of typical multi-day surveys usually do not exceed one week – most of the travel surveys are even limited to a single reporting day (e.g. most national microcensus data). Up to now, the available data bases have not allowed to perform an analysis of periodicity capturing weekly or even monthly rhythms of behaviour. The only known prior example of a multi-week travel diary is the 1971 Uppsala Household Travel Survey (see Hanson and Burnett, 1981; Hanson and Hanson 1981a, b) whose data is now only of historical interest and does not match the requirements of the current research.

With the implementation of a continuous six-week travel diary as part of the German research project Mobidrive, a current data set of long-term individual travel behaviour is now available for analysis (see Axhausen, Zimmermann, Schönfelder, Rindsfüser and Haupt, 2000). The survey was conducted in the cities of Halle/Saale and Karlsruhe in autumn 1999. A total of 317 persons over 6 years in 139 households participated in the main phase of the survey, after testing the survey instruments in a pre-test with a smaller sample in spring 1999 (44 persons). The paper-based travel-diary instrument was
supplemented by further survey elements covering the socio-demographic characteristics of the households and their members, the details of the households’ car fleet and transit season tickets owned and personal values as well as attitudes towards the different modes of transport.

The development of a stochastic model to detect rhythmic patterns of behaviour as presented here is based on the data generated within Mobidrive. The initial investigation of this unique data set reported here confirms the assumptions of a widely routinised character of travel behaviour and the impact of fixed temporal structures, such as working hours or leisure periods, over the course of the day and the week (see Axhausen et al., 2000; König, Schlich and Axhausen, 2000). Additionally, there are interesting findings concerning the different regional time-use structure in the western (i.e. Karlsruhe) and eastern part (i.e. Halle) of Germany.

2 CHOOSING A SUITABLE TOOL FOR ANALYSIS: SURVIVAL ANALYSIS

Analysing the character and the determinants of rhythmic patterns of individual travel behaviour requires further methodological development. In research settings treating aggregate count data such as traffic flows, the application of Time Series Analysis has been so far an effective tool of investigation (see Brockwell and Davis, 1991 for a basic introduction). This methodology seems to be inappropriate for the activity-based approach with individually assignable attributes of trips and activities. This is also because the interval data generated encompasses some special peculiarities which call for distinctive analysis approaches.

Rhythms can be characterised by their periodicity and their amplitudes, respectively. This paper focuses only on the periodicity of the behavioural rhythms shaped by the structure and duration of the intervals between two identical patterns of behaviour. The amplitudes of the rhythmic patterns may be chosen as simple or complex as needed – ranging from a one-dimensional description of activities by their purpose only, to a characterisation of complex sequences or daily patterns by several attributes. For the analysis of intervals as presented here, the durations between two identical activities – categorised by purpose only – were generated from the Mobidrive data set. In a further step, the duration data was related to additional socio-demographic attributes of the survey participants which are tested as co-variables in the proposed models.
2.1 **Special features of duration data and its analysis**

Linear regression models and related stochastic tools are usually applied to analyse distributions and the correlation of certain determinants with observed phenomenons. The application of those models using least-square estimation (LSE) is restricted in the field of duration data due to some model assumptions and the unique character of the empirical data (see Hosmer and Lemeshow, 1999). Empirical duration data as well as possible model estimates are limited to have positive values – negative time duration does not exist by definition!

Another constraint is the fact that for various reasons time intervals can be only observed and measured partially – the *censoring problem*. Limited observation periods usually effect the completeness of the measurements, since the relevant processes may have started before the beginning of the investigation or exceed beyond its end. Ordinary linear regression models cannot treat censored durations with values like “at least 5 months survival time”.

These constraints led to the development of *survival analysis* which is also used here as the suitable analysis and modelling frame for the MobiDrive interval data (see Kalbfleisch and Prentice, 1980; Cox, 1984; Kleinbaum, 1986 for basic references). The main aim of survival analysis is to represent the observed distribution of durations by the means of *hazard models*. Survival analysis is widely applied in other research fields such as biometry, mechanical engineering or market research. Survival analysis and especially hazard models are employed for forecasting the transition from one state to another – or in other words, the occurrence of a certain event following a period of time observed. In transportation research, hazard models have been employed since the end of the 1980s for a wide range of research questions – in the activity-based approach mainly for the conceptual support of scheduling tools and the general analysis of in- and out-of-home durations (Mannering and Hamed, 1990; Hamed and Mannering, 1993; Mannering, Murakami and Kim, 1992; Hamed, Kim and Mannering, 1993; Niemeier and Morita, 1994; Bhat, 1996a; 1996b; Ettema, Borgers and Timmermans, 1995; Reader and McNeill, 1999; Oh, 1999).

2.2 **Conceptual framework for the periodicity of travel behaviour**

The basic idea behind the application of hazard models for modelling cycles of time-use and travel behaviour is the concept of an *increase in demand* for carrying out an activity or a trip over time. The timing of the occurrence of an identified travel pattern (event) at a certain time *t* shall be expressed as a probability function *h*. The probability of occurrence *Pr* mainly depends on the time elapsed since
the last occurrence of the pattern ($T$; see figure below) and will presumably increase over time. Apart from the duration of the observed process external conditions set by the surrounding travel-environment will have an impact on the development of demand.

Undoubtedly, this concept is based on several theoretical constraints and simplistic behavioural assumptions. It is mainly grounded on the hypothesis that the assumed increase in demand probability is a monotone function of time. This presumes that individual behaviour is independent of any internal as well as external effects such as spontaneous modification of one’s own activity scheduling or the one of other individuals related with the actor. Besides, it remains questionable if the concept may be assigned to all activity categories in the same way – regardless whether the activity is discretionary or compulsory, constrained by fixed societal structures (e.g. opening hours) or flexible in space and time. Acknowledging these restrictions, this basic approach offers an initial explanatory tool for the explanation of the temporal form of the recurrence of activity performances over time.

2.3 Survival distributions: Mathematical basics

Survival analysis is a generic term for a group of models which characterise a probability distribution of the random variable $T$ – in other words, the time at which events occur is realised by some random
process causing a certain distribution of $T$. In our model, the event time $T$ defines the start time of an activity of a particular type respectively the end time of an interval between two identical patterns.

Before describing the actual models developed it seems useful to describe the mathematical functions representing the probability distribution of $T$ (Figure 2).

The **cumulative distribution function** $F(t)$ and its derivative $f(t)$ give the probability that an event will occur before or at least at some point in time $t$.

$$F(t) = \Pr[T < t]$$  \hspace{1cm} (1)

$$f(t) = dF(t)/d(t) = -dS(t)/dt = \lim_{dt \to 0} \frac{\Pr(t \leq T < t + dt)}{dt}$$  \hspace{1cm} (2)

Usually, what is more interesting than the question whether $T$ is less or equal to any value $t$ is surviving a process beyond a certain point – such as the surviving beyond the end of the observation period. The **survival function** $S(t)$ expresses the related probability, i.e. surviving beyond $t$. As $S(t)$ is a probability, the function is limited to values between 0 and 1 and must be non-negative by definition. Besides, $S(0) = 1$. The function takes diverse shapes according to the character of the processes observed. Along with the restrictions mentioned, it is decreasing in most cases.

$$S(t) = \Pr[T \geq t] = 1 - F(t)$$  \hspace{1cm} (3)

The most common function representing the distribution of durations is the **hazard function** $h(t)$ which is essential for the further modelling process. All three functions described above can be associated with the hazard function. It gives the probability or the direct risk that the occurrence of an event can be expected in a (small) interval between $t$ and $dt$ - provided that the event has not occurred until this point in time. Thus, only individuals (processes) are considered which belong to the actual risk set, i.e. which survived until the beginning of the mentioned time interval. In contrast to the probability density function – note the similarity of terms – the hazard function represents a conditional density which follows the restrictions of the risk set.

$$h(t) = f(t)/S(t) = \lim_{dt \to 0} \frac{\Pr(t \leq T < t + dt|T \geq t)}{dt}$$  \hspace{1cm} (4)

Due to its definition, the interpretation of the hazard rate (as part of the model output) requires some care. Hazard rates actually represent latent intensity variables of transition from one state to another.
rather than probabilities in a narrower sense (see Schneider, 1991): The higher the value, the quicker the transition from state A to state B takes place on average.

Figure 2  Survival analysis: Main distributional representations


3  INITIAL DATA ANALYSIS BY THE NON-PARAMETRIC APPROACH

The investigation of durations should start with a descriptive analysis of the spells by a non-parametric duration model. This analysis aims to estimate the survival rates and their corresponding hazard rates – based on the observed intervals between two identical patterns of behaviour. The empirical survival function is given by

$$\hat{S}(t) = \frac{\text{Number of Observations} \geq t}{n} \quad t \geq 0$$  \hspace{1cm} (5)

Empirical survival rates are usually calculated by the Kaplan-Meier estimator (KM) – also called product limit estimator – which decomposes the survival of processes to any point in time into a series
of steps. Those steps are either defined by the observed survival and censored times or by plausibly
determined intervals. The latter case is known as life-table method and is reasonable to use if many
exactly measured durations, i.e. many unique event times, are to be analysed. Considering the
Mobidrive database with more than 50,000 activities as well as the same amount of intervals between
them, this method was chosen for analysis. The main advantage compared to the usual Kaplan-Meier
estimator is the fact that fewer parameters are estimated which eases the interpretation of the results.

Fundamental principle of the life-table method is the grouping of the measured events and censored
times into categories or intervals. This is combined with the assumption of a uniform distribution of
censored times over the defined interval. The average size of the risk sets in the interval becomes
\( n - \left( \frac{c}{2} \right) \), with \( n \) equals the number of subjects at risk of dying in the interval and \( c \) equals the
durations censored within the interval. To simplify matters, no censored durations of the intervals were
considered in this analysis. Selecting the width of the intervals is an arbitrary process which usually
requires some testing – in this case a width of 24 hours was believed to be the suitable interval.
This initial analysis offers first interesting insights into the structure of activity scheduling and performance over the six-week reporting period of the Mobidrive survey. The intervals were examined for the twenty-three activity categories coded – at this time, we just present a subset from the overall...
results focusing on the periodicity of leisure and shopping activities. In contrast to the obligatory categories such as work or school which should have significant regularity over time the periodicity of discretionary activities over prolonged periods has not been explored yet.

Our results show a clear rhythmic character of activity performance with characteristic shapes of the curves. Interestingly, leisure activities – categorised in detail – tend to follow a fairly stable rhythm. This is obvious with the categories active sports and club meeting with a relative maximum of transition intensities (hazard rates) at around one week. The other activity categories show similar tendencies with moderate significance, though.

4 PARAMETRIC DURATION MODELS

The descriptive analysis presented in the last section does not answer the question which determinants other than the observed durations actually effect the periodicity in time-use and travel behaviour. Addressing this question, an initial semi-parametric as well as a fully-parametric hazard model will be presented.

4.1 Proportional hazard models

Hazard models not only allow to incorporate measured durations as essential variables for the probability of event occurrence, they are also able to treat duration-independent determinants such as socio-demographic attributes or personal constraints of the travellers. Those determinants play an important role for the development of a more realistic model expressing the periodicity of travel-behaviour as a result of a complex structure of personal and environmental factors. In the following, those factors will be investigated more closely by proportional and parametric hazard models.

In general, proportional hazard models treat additional explanatory variables as a function of a multi-dimensional vector $X$ which has multiplicative effects on an underlying (baseline) hazard. For non-parametric basic models it is assumed that all explaining co-variables equal zero and do not account for any effects. Thus, the hazard function presented in Equation 6 is a product of two hazard functions – with $h_0(t)$ as a function of survival times whereas $g_0(t)$ gives the change of the hazard function caused by subject co-variables.
Analysing the rhythms of travel using survival analysis

August 2000

\[ h(t|X) = h_0(t)g_0(X) = h_0(t)\exp(\beta X) \]

with \( X \) = vector of co-variables
\( \beta \) = vector of parameters
\( h_0(t) \) = baseline hazard

Calculating the logarithm of both sides of the equation, one obtains

\[ \log h_i(t) = \alpha(t) + \beta_1x_i + \ldots + \beta_kx_k \]

This equation is important for further distributional baseline assumptions as described later.

The characteristics of the hazard function change according to the values of the co-variables, given that the ratio of the hazard function remains stable over time (i.e. assumption of constant hazards; \( \approx h_i/h_j \)). This means that the hazard rate of an individual \( i \) is a fixed proportion (ratio) of an other individual \( j \) with different personal attributes. The hazard ratio is given by

\[ \frac{h_i(t)}{h_j(t)} = \left\{ \exp \beta_1(x_{i1} - x_{j1}) + \ldots + \beta_k(x_{ik} - x_{jk}) \right\} \]

There is a wide range of approaches to proportional hazard models which can be distinguished by their distributional assumptions for the baseline hazard. The shape of hazard functions for temporal processes may have different characters such as monotone-increasing, U-form shape, monotone-decreasing or constant. All those different shapes are caused by the distributional assumptions for durations, and one is free to choose a specific distribution meeting one’s own theoretical assumptions. For example, monotone-increasing hazard functions (Weibull distribution) may represent processes whose end is going to be more probable the longer they last, whereas constant hazards (exponential distribution) cover processes with no duration dependence.

Initially, Cox proportional hazard models were estimated which are probably the most widely used approach in the field of survival analysis (see Cox, 1972). Its popularity can be explained by the fact that the method does not require a choice of any particular probability distribution to represent event times. Several related applications in transport research also followed this methodology in the last years (see Hensher and Mannering, 1994; Sueyoshi, 1993).

Due to the omission of assumptions concerning the baseline hazard the model is called semi-parametric. Cox models solely focus on the estimation of the parameter vector \( X \) (functional form). In the literature, there is agreement that leaving out the distributional assumptions for the underlying hazard protects one from drawing wrong conclusions about the actual distribution of \( T \). Therefore, the
Cox models are regarded as more robust than the fully-parametric approaches. Another advantage is the possibility to integrate time-varying determinants which take different values over the course of the observation period (e.g. growing experience due to cognitive learning processes or weather effects). Finally, discrete as well as continuous measured durations may be considered.

Realising the advantages of the Cox approach there are downsides, though, which are mainly associated with the neglecting the *heterogeneity effects* of the respondents. Neglecting unobserved heterogeneity could lead to biases in the model estimation – especially to decreasing hazard rates even if the real hazard is not decreasing at all for any individual. On the other hand, considering heterogeneity effects requires a further distributional assumption for the heterogeneity itself which is linked with additional serious uncertainties. Most authors argue though that there is enough room for randomness in the relationship between the unobserved hazard and the observed event time.

Cox models are based in the principle of *partial likelihood* (see e.g. Kleinbaum, 1996, pp. 99 for detail). Partial likelihood generally ignores the first part of equation 6 and treats the second as if it is an ordinary likelihood function. The fact that the first term (*baseline hazard*) could contain time dependent information about the parameter $\beta$ does not automatically lead to severe biases.

### 4.2 Exemplary model estimation with the Cox approach

The wide range of available activity categories in Mobidrive offers opportunities for a detailed analysis of rhythmic structures bearing in mind that the periodicity of activity performance differs a lot from one category to the other. Hence, a further aggregation such as mandatory versus discretionary activities is not made here. At this point in the analysis two exemplary Cox models of interval durations are presented for the activities *daily shopping* and *active sports*. The models were developed separately for each case study city. The estimated survival functions are presented in the following figure. Please note the similarity to the empirical curves in Figure 3.
Figure 4  Estimated survival rates at the means of co-variables by Cox duration model (intervals exceeding fourteen days are not displayed)

a) Karlsruhe: Daily shopping (n = 2024)  
b) Halle: Daily Shopping (n = 1737)

Before discussing the selected model results, it seems appropriate to indicate some specific statistical aspects of hazard modelling. First of all, the proportionality assumption for the baseline hazard should be tested. There is a range of approaches testing for proportionality (see Kleinbaum, 1996, pp 129) – for the evaluation of our model a standard graphical method was applied. The log-log survival curves for the categories of the chosen co-variables should be parallel if the hazards are proportional. All models presented in the following meet the proportionality assumptions – at least by an initial graphical examination of the estimated curves for each co-variable separately.

The top of the following table provides the parameter estimates and the calculation of the risk ratios (see below). Partial likelihood models do not provide any estimates of the intercept which is part of the logged baseline function (see Equation 7). Of primary interest for the interpretation is the risk or hazard ratio which is simply calculated as $e^\beta$ for Cox models. The risk ratio is a quantitative indicator for the multiplier effects of the co-variables of one individual towards another. For indicator variables
such as sex the risk ratio is simply the ratio of the estimated hazard with a value of 1 to the estimated hazard for those with a value of 0 (controlling for other co-variables) (see Allison 1995, p. 117). For example, the risk ratio for the Karlsruhe respondents reporting more than 30 work hours per week in the model of short-term shopping is 0.66. This means that their hazard rate is only 66% of all other respondents in the sample which results in a lower frequency of activity performance over time. For continuous variables, a reasonable estimate is found by subtracting 1 from the risk ratio and multiplying the result by 100. This gives the estimated percentage change of the hazard rate per each unit increase in the co-variable. For the co-variable number of cars in the household in the same model, there is a risk ratio of 0.85 which leads to an approximately 15% lower hazard rate for each additional car in the household.

Table 1 Model estimation with semi-parametric Cox model; Activities: Daily shopping and Active sports – Positive tests of proportionality

<table>
<thead>
<tr>
<th></th>
<th>Karlsruhe</th>
<th></th>
<th>Halle</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Daily shopping</td>
<td>Active sports</td>
<td>Daily shopping</td>
<td>Active sports</td>
</tr>
<tr>
<td>β</td>
<td>Risk Ratio</td>
<td>β</td>
<td>Risk Ratio</td>
<td>β</td>
</tr>
<tr>
<td>Male</td>
<td>-0.03</td>
<td>1.03</td>
<td>0.39</td>
<td>1.48*</td>
</tr>
<tr>
<td>Age</td>
<td>0.03*</td>
<td>1.03</td>
<td>0.01</td>
<td>1.01</td>
</tr>
<tr>
<td>Age (squared)</td>
<td>-0.00</td>
<td>1.00</td>
<td>-0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>License holding</td>
<td>-0.17</td>
<td>0.85*</td>
<td>0.50</td>
<td>1.65*</td>
</tr>
<tr>
<td>Club member</td>
<td>-0.01</td>
<td>1.00</td>
<td>0.62</td>
<td>1.86*</td>
</tr>
<tr>
<td>Married</td>
<td>0.03</td>
<td>1.03</td>
<td>0.01</td>
<td>1.01</td>
</tr>
<tr>
<td>Parent</td>
<td>0.24</td>
<td>1.27*</td>
<td>-0.26</td>
<td>0.77</td>
</tr>
<tr>
<td><strong>Household related</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of HH members</td>
<td>-0.09</td>
<td>0.92*</td>
<td>0.08</td>
<td>1.08</td>
</tr>
<tr>
<td>Number of cars</td>
<td>-0.16</td>
<td>0.85*</td>
<td>0.03</td>
<td>1.02</td>
</tr>
<tr>
<td>Net income: &gt; 2000DM / HH member</td>
<td><strong>0.31</strong></td>
<td>1.36*</td>
<td>-0.03</td>
<td>0.97</td>
</tr>
<tr>
<td><strong>Occupation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-time &gt; 30 h / week</td>
<td><strong>-0.48</strong></td>
<td>0.62*</td>
<td><strong>-0.28</strong></td>
<td>0.76*</td>
</tr>
<tr>
<td>N</td>
<td>2024</td>
<td>672</td>
<td>1737</td>
<td>301</td>
</tr>
<tr>
<td>LogLikelihood(0)</td>
<td>26781</td>
<td>7417</td>
<td>22454</td>
<td>2843</td>
</tr>
<tr>
<td>LogLikelihood(β)</td>
<td>26566</td>
<td>7322</td>
<td>22381</td>
<td>2755</td>
</tr>
<tr>
<td>ChiSquare (DF=11)</td>
<td>215</td>
<td>95</td>
<td>73</td>
<td>88</td>
</tr>
<tr>
<td>R²</td>
<td>0.101</td>
<td>0.133</td>
<td>0.041</td>
<td>0.252</td>
</tr>
</tbody>
</table>

* Statistically significant at 95%-confidence interval
Interpreting the model results, there are some findings worth mentioning: Firstly, it can be assumed that the chosen co-variables just have moderate explanatory impact on the estimated model output (see the low $R^2$). Analogous for both cities, the co-variates in the model for *active sports* activities are more strongly associated with the dependent variable than in the *shopping* activity model – represented by a higher $R^2$-value. The selection of co-variables was made initially by some ad-hoc assumptions concerning the determinants of travel behaviour as well as by the requirement to meet the proportionality assumptions of the model. They should be further specified in future work. Related tests will focus on the statistical significance of single explanatory variables as well as the interactions with others and will also be based on a more detailed conceptual approach to rhythmic patterns.

It could be found that there is a significant difference in the determinants for the temporal behaviour between the samples in the two case study cities – resulting from the actual socio-demographic structure of the sample on the one hand and from probable behavioural heterogeneity on the other. One additional, though more quantitative explanation is the fact that the Karlsruhe respondents reported a 20% higher trip rate than their Halle opposites. In other words, a separate Cox model for shopping activities which exclusively considered a dummy variable for the case study cities would automatically lead to a significantly higher hazard rate for the city of Karlsruhe.

As indicated, the estimated parameters especially differ in the model for *daily shopping* activities. While the Karlsruhe sample shows shorter intervals for automobile travellers with higher vehicle availability and license holding the Halle sample does not. Here, longer intervals for the male respondents could be found. A common result for both cities is the statistical significance of full-time employment which yields longer intervals between shopping activities and the impact of a bigger household income which leads in contrast to a higher activity frequency.

The co-variables sex, license holding and club membership are of special interest in the hazard model for the activity category *active sports*. Especially men and – as expected – club members show shorter intervals. Interestingly, car license holders prove to have an amazing sports activity frequency. See the hazard rate more than two times larger than all other respondents. Also significant are the risk ratios of the full-time employed participants (more than 30 work hours per week) which are considerably below those of all other respondents.

### 4.3 Fully-parametric hazard model with a Weibull baseline distribution

Fully parametric models which are based on an explicit distributional assumption for the *baseline hazard* may be realised either by proportional hazard or *accelerated failure time* (AFT) models. The
latter class of models describes the relationship between the survival function of any two individuals or processes so that

$$S_i(t) = S_j(\phi_{ij} t)$$ (9)

with $\phi_{ij}$ = constant specific to the pair $(i, j)$.

Linking AFT models to the proportional hazard assumption (Equation 6), they can be defined as

$$S(t|Z) = S_0(t \exp(\beta Z))$$ (10)

with the corresponding hazard function

$$h(t|Z) = h_0(t \exp(\beta Z)) \exp(\beta Z)$$ (11)

with $Z$ = vector of co-variables

According to recent work which recommend the Weibull distribution as one major descriptor for the relationship of behaviour and duration (Hamed and Mannering, 1993; Hensher and Mannering, 1994; Bhat, 1996a, b), this distributional assumption was chosen for our purposes, too. The Weibull model – based on a Weibull distribution for $T$, conditional on the co-variables – also showed a larger in the loglikelihood values compared with other models based on different distributional assumptions (exponential / log-logistic).

Weibull models comprise a simple survival function with

$$S(t) = \exp\left\{-\left[\frac{t}{\lambda_e} \right]^\beta\right\}$$ (12)

Since Weibull models belong to the group of AFT models as well as to the proportional hazard model class its estimates may be easily translated into relative survival ratios when suitable transformed before. This is in line with the semi-parametric Cox approach presented earlier. The Weibull distribution is a modification of the exponential with the constant hazard $h(t) = \lambda$ – representing duration independence of event occurrences. The Weibull model itself does not share this restrictive presumption, it rather leads to a monotone character of the hazard with the density function given by
Analyzing the rhythms of travel using survival analysis

August 2000

16

(13) \[ f(t) = \lambda P(\lambda t)^{p-1} \exp\left[-(\lambda t)^p\right] \]

with \( P \) = Weibull parameter

and the hazard function

(14) \[ h(t) = \lambda P(\lambda t)^{p-1} \]

The direction of the monotony is dependent on the value of the Weibull parameter with a monotone increasing hazard if \( P > 1 \) and a monotone decreasing hazard if \( P < 1 \). The Weibull model is reduced to the exponential form if \( P \) equals 1. It still inherits restrictive theoretical assumptions due to the monotone character of the hazard (see Hensher and Mannering, 1994 and the final discussion) but eventually allows some duration dependence matching our concept of increase-in-demand.
Table 2  Estimates by fully-parametric approach with co-variable effects; Activities: Daily shopping and Active sports

<table>
<thead>
<tr>
<th>Aktivitäten</th>
<th>Karlsruhe</th>
<th>Halle</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Daily shopping</td>
<td>Active sports</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>Surv. Ratio</td>
</tr>
<tr>
<td>Baseline parameters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda$</td>
<td>0.48 n.a.</td>
<td>0.32 n.a.</td>
</tr>
<tr>
<td>$P = 1/\sigma$</td>
<td>0.80 n.a.</td>
<td>1.04 n.a.</td>
</tr>
<tr>
<td>Intercept</td>
<td>1.00 n.a. *</td>
<td>2.26 n.a. *</td>
</tr>
<tr>
<td>Personal attributes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>-0.04 0.96</td>
<td>-0.39 0.68 *</td>
</tr>
<tr>
<td>Age</td>
<td>-0.03 0.97 *</td>
<td>-0.01 0.99</td>
</tr>
<tr>
<td>Age (squared)</td>
<td>0.00 1.00 *</td>
<td>0.00 1.00</td>
</tr>
<tr>
<td>License holding</td>
<td>0.20 1.22 *</td>
<td>-0.46 0.63</td>
</tr>
<tr>
<td>Club member</td>
<td>-0.01 0.99</td>
<td>-0.63 0.53 *</td>
</tr>
<tr>
<td>Married</td>
<td>-0.02 0.98</td>
<td>-0.01 0.99</td>
</tr>
<tr>
<td>Parent</td>
<td>-0.29 0.75 *</td>
<td>0.25 1.28</td>
</tr>
<tr>
<td>Household related</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of HH members</td>
<td>0.11 1.12 *</td>
<td>-0.08 0.92</td>
</tr>
<tr>
<td>Number of cars</td>
<td>0.19 0.12 *</td>
<td>-0.01 0.99</td>
</tr>
<tr>
<td>Net income: $&gt; 2000DM / HH member</td>
<td>-0.36 0.70 *</td>
<td>0.10 1.10</td>
</tr>
<tr>
<td>Occupation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-time $&gt; 30$ h / week</td>
<td>0.57 1.77 *</td>
<td>0.26 1.30 *</td>
</tr>
<tr>
<td>N</td>
<td>2024</td>
<td>672</td>
</tr>
<tr>
<td>LogLikelihood(0)</td>
<td>-3807</td>
<td>-1078</td>
</tr>
<tr>
<td>LogLikelihood($\beta$)</td>
<td>-3707</td>
<td>-1025</td>
</tr>
<tr>
<td>ChiSquare (DF=11)</td>
<td>200</td>
<td>106</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.09</td>
<td>0.15</td>
</tr>
</tbody>
</table>

* Statistically significant at 95%-confidence interval
** Statistically significant at 90%-confidence interval

Considering the differences between the Weibull fully-parametric and the Cox model, the estimates are similar to the results generated by the semi-parametric approach. Statistical significance is given mainly for the variables income and occupation in the daily shopping and the variables sex, car license, club membership and again occupation in the active sports model. In contrast to the estimates generated by the Cox approach positive coefficients represent longer survival times for individuals whereas negative coefficients are associated with shorter times. In other words, the negative sign indicates a higher transition rate or shorter intervals between two identical patterns of behaviour, respectively. This is because the hazard is given by $\lambda = \exp(-\beta Z)$ in fully-parametric approaches. The estimates may be again transformed into more informative figures by some simple calculations: the
multiplicative effect of the chosen indicator (dummy) co-variables may be found by taking $e^\beta$ which yields the estimated ratio of the expected (mean) survival times for the two groups (Allison 1995, p. 65). Note the obvious difference to the risk ratio estimated in the Cox approach. For count variables, an analogous indicator is obtained by $100(e^\beta-1)$ which provides the percent increase in the expected survival time for each one-unit increase in the variable.

In contrast to these findings which are parallel to those of the Cox model, the Weibull distribution causes some unexpected, though interesting effects concerning the baseline hazard. For the daily shopping model it produces a decreasing hazard ($P < 1$) which tends to be as good as constant soon after day one (Figure 5).

Figure 5 Baseline hazard rates at the means of co-variables by fully-parametric duration model (intervals exceeding fourteen days are not displayed)

a) Daily shopping (Karlsruhe: n = 2024; Halle: n = 1737)

This is contradictory to our intuitive expectations concerning an increase in demand between two behavioural patterns – at least for this activity category. By further initial inspection of the model estimates for the other activity categories not presented here, it can be found that all discretionary activities with a flexible rather than a regular structure of performance show the same tendency (e.g. long-term shopping). It may be assumed that the nearly constant hazard indicates a regular demand over time with no significant rhythmic pattern – at least for the pooled data of the whole sample. For frequent respectively regular activities with considerably shorter intervals, the Weibull model causes increasing underlying hazards in contrast.
5 DISCUSSION AND METHODOLOGICAL OUTLOOK

Applying hazard models to the analysis of the periodicity in travel behaviour may supplement the activity-based methodology by the detection of cyclical temporal structures. Considering the initial character of our stochastic approach there are several conceptual and modelling issues which have to be worked out more detailed in future work. Conceptually, the focus of work will lie on the question whether the increase-in-demand is a monotonous phenomenon and if it requires a variable treatment of the different activity categories. The results obtained by the Weibull model indicate that the latter question deserves special care. In addition to that, it still remains open if time-use and travel behaviour may be assumed to be homogenous across the whole sample.

Similar profound problems are to be solved in future work when addressing the modelling issues in detail. While the empirical inspection of the interval duration data by non-parametric models turned out as a straightforward analysis of the inherent temporal structure, the development of parametric models requires

- a purposeful selection of co-variables by means of detailed significance tests
- a further discussion about reasonable distributional assumptions for the restrictive fully-parametric approaches
- the inclusion of censored times which were omitted so far and
- the integration of heterogeneity effects caused by the behavioural differences within the sample, the effects of repeated events or state dependency.

Finally, the initial concept may easily be complemented by additional aspects such as the consideration of more complex behavioural patterns or the distinction of different kinds of events (competing risks). The latter aspect will have implications for the choice of the modelling approach.

Summing up the initial findings of the models developed here, the estimations confirm most general assumptions concerning the determinants of travel such as the socio-demographic attributes of the travellers and their consumption as well as leisure habits and performance. For the first time though, those determinants can be associated with quantifiable information about the periodicity of everyday’s life. The results will especially provide computational process modelling (CPM) or scheduling tools with essential information concerning the temporal allocation of recurrent behavioural patterns.
Acknowledgements

The authors would like to thank the Mobidrive project partners T Haupt and A Zimmermann (PTV AG, Karlsruhe), KJ Beckmann and G Rindsfüsser (Institut für Stadtbauwesen, RWTH Aachen) as well as especially our IVT colleagues A König and R Schlich for their support preparing this paper.
References


Hanson, S. and Hanson, P. (1981a) The travel-activity patterns of urban residents: Dimensions and relationships to sociodemographic characteristics, *Economic Geography*, 57(4), 332-347.

Hanson, S. and Hanson, P. (1981b) The impact of married women’s employment on household travel patterns: A Swedish example, *Transportation*, 10(2), 165-183.


Eine vollständige Liste der Berichte kann vom Institut angefordert werden:

A complete catalogue of the papers can be obtained from:

IVT
ETH Hönggerberg
CH - 8093 Zürich

Telephon: +41 1 633 31 05
Telefax: +41 1 633 10 57
E-Mail: hotz@ivt.baug.ethz.ch
WWW: www.ivt.baug.ethz.ch

Der Katalog kann auch abgerufen werden unter:

The catalogue can also be obtained from:

http://www.ivt.baug.ethz.ch/veroeffent_arbeitsbericht.html