Institute of Transport Studies
Graduate School of Business
The University of Sydney
NSW 2006

Working Paper
ITS-WP-94-2

THE TIMING OF CHANGE FOR AUTOMOBILE TRANSACTIONS: A COMPETING RISK MULTISPELL SPECIFICATON

David A. Hensher

February 1994

A paper for presentation at the International Conference on Travel Behaviour, Valle Nevado, Santiago, Chile, 13-16 June 1994.

NUMBER: Working Paper ITS-WP-94-2

TITLE: The Timing of Change for Automobile Transactions: A Competing Risk Multispell Specification

ABSTRACT:

With the increasing number of panel data sets available in transport, the opportunity exists for the study of the time frame a household uses in making transport decisions. Panels collect data at regular points in time, and record information related to occurrences over the period since the last wave. Some panels record the precise time of an event during a panel wave. The opportunity to record event histories complete with identification of the states associated with a phenomenon such as automobile ownership together with the duration spent in each state provides powerful data for modelling the 'when' component of change through time. The ability to trace the timing of change and to model it will give transportation planners one missing element of forecasting - the timing of change. This paper develops a number of competing risk multispell models to obtain insights into the time spent in each of three states of automobile transactions (no change, replace used with used vehicle, and replace used with new vehicle), the factors which affect the probability of leaving a state, the probability of staying with a state, the effect of past history on current behaviour, and whether the population segments into distinct groups with different change probabilities. A data set of 12 years of annual observations (1974-1985) of a sample of Sydney households is used to illustrate the application of event history models.

AUTHOR: David A. Hensher

CONTACT: Institute of Transport Studies
Graduate School of Business
University of Sydney NSW 2006
Australia

Telephone: + 61 2 550 8631
Facsimile: + 61 2 550 4013
E-mail: davidh@gsb.su.oz.ai

DATE: February 1994

1. INTRODUCTION

A knowledge of the length of time that a household keeps an automobile, of the timing of a change in the fleet, of the type of ensuing transaction (e.g. replace a used vehicle with a new vehicle, replace a used vehicle with a used vehicle), and of the influences on the time frame and transaction type are important issues in forecasting and hence transport planning. They are of interest to many planning agencies. For example, government authorities need to identify the time it takes for a household to dispose of older less fuel efficient vehicles under various pricing and structural contexts, since this lag is important in the development of forecasts of the impact of policies designed to meet emission reduction targets. An understanding of the reasoning behind the turnover or lack of turnover of vehicles in households is also important to automobile manufacturers; it provides some guidance on how the demand for new vehicles is changing over time and what the likely influences on the timing and speed of change might be. Scrappage rates of vehicle classes can be developed out of this framework.

The growing availability of transport panel data has expanded the opportunities to develop models to identify the temporal relationships (i.e. timing and duration) between automobile acquisitions and disposals, and the influences on the timing and duration of automobile ownership. Transport panels are typically of a limited life with repeat waves every 6 or 12 months over 4 to 6 years (see Raimond and Hensher 1993 for a review, also Axhausen 1992, Golob and Golob 1989, Murakami and Watterson 1990, and van Wissen and Meurs 1989). Some panels identify the precise date of key events (e.g. vehicle replacement), enabling the richness of the timing and duration of events to be identified. The Sydney automobile panel offers this richness. In each of the four waves of the Sydney automobile panel (Hensher et al. 1992), spaced 12 months apart, details were sought on the annual profile of the automobile fleet composition and utilisation, and the socioeconomic description of each household member. Data on a selected set of items however were obtained over a longer period using a retrospective recall strategy in order to identify the initial conditions for the panel. This data is sufficiently rich to allow us to develop a 12 year panel (1974-1985) on a limited number of socioeconomic and automobile variables.

The aim of this paper is to utilise the restricted 12 year panel to investigate alternative ways of modelling the automobile transactions decision for a sample of 200 households in Sydney. Three states or a maximum possible 9 changes of state (i.e. transactions or transitions) are investigated. The states are no change (S1), replace a used vehicle with a used vehicle (S2), and replace a used vehicle with a new vehicle (S3). Other feasible states infrequently observed in the data set (and excluded herein) are: dispose of a vehicle (S4), acquire a used vehicle (S5) and acquire a new vehicle (S6). The modelling of movement from multiple origin states to multiple destination states and the desire to preserve the distinction between the OD transitions, as well as permitting repeated presence in each transition type, adds considerable complexity although significant realism over the majority of transportation applications of the tools outlined in this paper.

The paper is organised as follows. The next section provides an overview of methods suitable for modelling event histories, followed by a discussion of the context in which the models will be implemented. A set of estimated models are then reported and interpreted, followed by a number of main conclusions.

2. AN EVENT HISTORY APPROACH TO STUDYING THE TIMING AND DURATION OF CHANGE

Real choice opportunities and ensuing decisions are inherently dynamic. When we observe a choice there is a history of events that have preceded the current outcome. By deduction, the role of the forces that have shaped the timing and duration of an event history are likely to have an important role to play in the continuing evolution of decision making. If the analyst can ‘capture’ the structure of an event history through some formal quantitative procedures, the ability to predict the timing of future change and hence the amount of time spent in a particular event state is likely to be substantially enhanced. Given the inherent uncertainties in any assessment of future event paths, all outcomes are probabilistic.

An event history can be profiled in a timeline (Figure 1). In principle the time dimension can be graduated to any level of refinement; in practice however the recording of events is often truncated to a number of finite discrete time periods. When the discretisation is sufficiently fine such that a ratio-scale treatment is feasible in the time dimension, a continuous time specification of an event history model is possible. This is essential if we are to study the duration of events.

Event histories can be characterised as a time-sequenced set of events. For each unit of analysis, event histories provide information about the exact duration until a state transition as well as the occurrence and sequence of events. For example, such data can provide information on the amount of time a household held a particular vehicle, the exact dates of acquisition and disposal, and the nature of the transaction at the time of disposal (eg.

replacement or disposal only). This information can be identified for each vehicle in a given period of time.

A duration model in its statistical form is referred to as a hazard function. Formally, the hazard function can be expressed in terms of a cumulative distribution function, F(t), and a corresponding density function, f(t). The cumulative distribution is written as,

$$F(t) = \text{Prob}[T < t] \tag{1}$$

where Prob denotes the probability, T is a random continuous time variable, and t is some specified time. Equation (1) for example, identifies the probability of replacing a vehicle before some transpired time (assuming no left censoring). The corresponding density function is

$$f(t) = dF(t)/dt \tag{2}$$

and the hazard function is,

$$h(t) = f(t)/[1 - F(t)] \tag{3}$$

where h(t) is the conditional probability that an event will occur between time t and t+dt given that the event has not occurred up to time t:

$$\text{Prob}\ (T_o \geq t+1 |\ T_o \geq t) \tag{4}$$

Information relating to duration dependence, as derived from the first derivative of the hazard function with respect to time (ie. its slope) provides insights into the duration process being modelled. Plotting the hazard function against time gives important empirical information for the parameterisation of the baseline hazard (Hensher and Mannering 1994). The probability of ending a duration or spell in a particular state may be dependent on the length of the duration. There may also be important determinants of duration (eg. socioeconomic characteristics) that should be included in the modelling approach. These covariates are included in hazard-based models using two alternative methods; *proportional hazards* and *accelerated lifetime*.

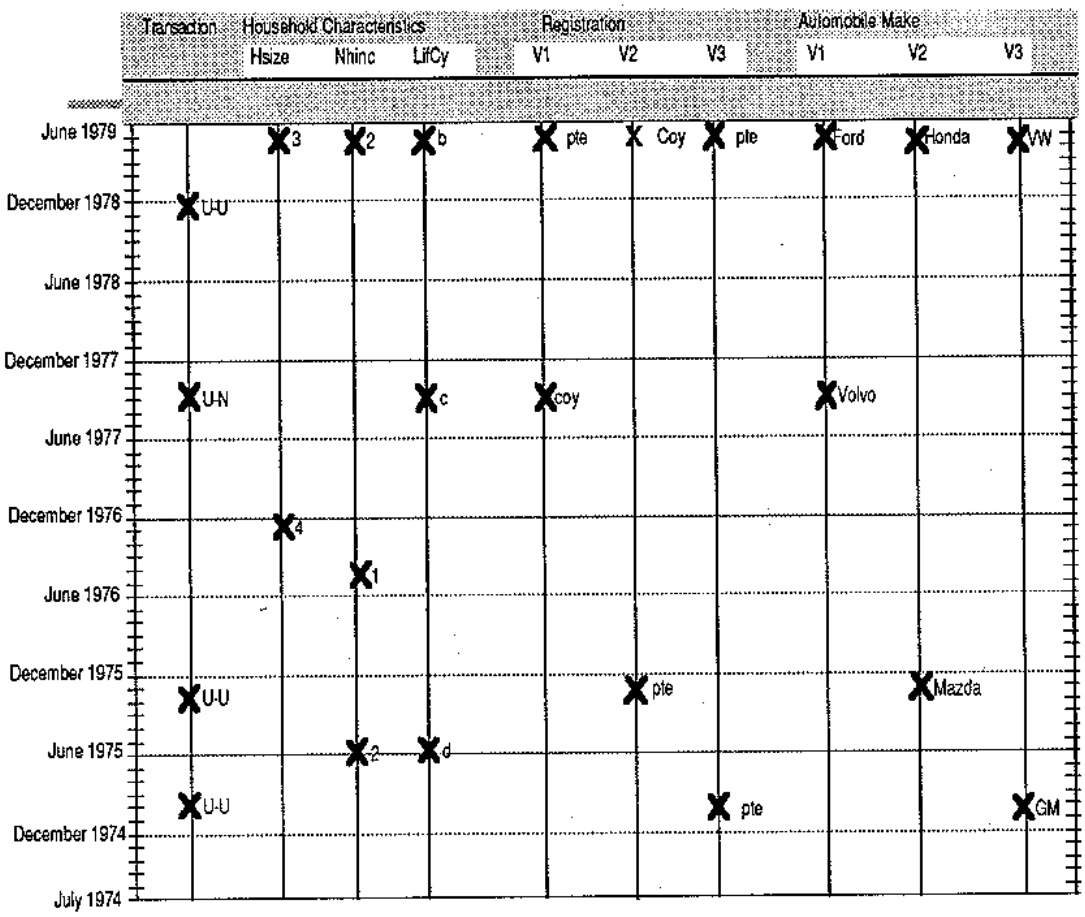

Figure 1 Timeline for Automobile Transactions

Proportional hazards models operate on the assumption that covariates act multiplicatively on some underlying or baseline hazard function. The proportionality is due to the decomposition of the hazard rate into one term dependent upon time, and another dependent only on the covariates (Prentice and Gloeckler 1978). To accommodate time varying covariates we assume that they are well approximated by their mean over the interval. This gives a clue to the interval size given the particular application (Hensher and Raimond 1992). A relatively general form of the hazard is specified as:

$$h_o(t) = \lambda_b(t) \exp (z_o(t)\beta) \tag{5}$$

where $\lambda_b(t)$ is an arbitrary baseline hazard and $\exp (z_o(t)\beta)$ is the parametric component including time varying covariates associated with an origin state *o*. A discrete set of time intervals are observed. The conditional probability rule in (4) translates into the following function, given (5):

$$\text{Prob}\ (T_o \geq t+1 \mid T_o \geq t) = \exp\ (\ -\exp\ (\gamma\ (t) + z_o(t)\ \beta)) \qquad (6)$$

where $\gamma\ (t) = 1n\ (\int_t^{t+1} \lambda_b(u)du)$. u is any function in terms of time. The model allows for a continuous 'failure' time T_o and (right) censoring c_o, but with observation taking place only at t_o, t = 0,1,2,..., *J*-1, or in the final interval (*J*, ∞). If the baseline hazard is assumed to be well approximated by its mean over the time interval, it is completely captured by the single term $\gamma\ (t)$. Left censoring may exist if an event was well under way when the panel commenced. Right censoring exists since the endpoint of the last episode of an individual cannot be observed. We allow for right censoring in model estimation.

An alternative approach for incorporating covariates in hazard-based models is the accelerated lifetime model. This model assumes that the covariates rescale time directly (i.e. accelerate time). Assuming that the covariates act in the form exp(βZ), as was the case for the proportional hazards model, the accelerated lifetime model can be written in terms of hazard functions as:

$$h(t|Z) = h_b[t\exp(\beta Z)]\exp(\beta Z) \qquad (7)$$

Accelerated lifetime models, along with proportional hazards (PH) models, enjoy wide-spread use (see Kalbfleisch and Prentice, 1980). The selection of accelerated lifetime or proportional hazards models is often determined on the basis of distributional assumptions (i.e. the assumed distribution of durations - Weibull, normal, gamma etc.). We concentrate on the PH model for the rest of the paper.

2.1 Competing Risks and Multispell Models

The dominating emphasis in empirical analysis of event history data (particularly in transportation, but also in economics (eg. Lancaster 1979) and marketing (eg. DuWors and Haines, 1990)) involves the study of a single initial or origin state, a single final or destination state, and a single period of time between successive events, often referred to as a single episode or spell. In the marketing literature it is called a non-repeated event. An example of the singular dimensionality would be studying the time before a traveller switches from a free public route to a tolled private route (Hensher and Raimond 1992). Multistate (or competing risks) and multispell situations are common in transportation, but they impose substantial complexity on the estimation of models. The combination of complexity and the general absence of packaged software for multistate and multispell models has limited applications, despite the realism. LIMDEP (Econometric Software, 1993), for example, currently only handles a single initial state, a single destination state and a single spell. Hamed and Mannering (1992) and Paselk and Mannering (1992) are two important transport applications using a single state single spell framework.

The application of interest herein involves three origin states (no change - O1, replace used with used vehicle - O2, and replace used with new vehicle - O3) and three destination states (the same three states), to give 9 possible OD states or transitions. In reality the only 5 transitions likely to be observed are O1-D1, O1-D2, O1-D3, O2-D1 and O3-D1. Furthermore we want to preserve the distinction between each pair of states and allow for repeated transitions from one state to another or repeated occurrence of events.

In the past however, many researchers have assumed that a competing risks model with *n* possible outcomes, had a likelihood function that could be separated into *n* distinct pieces. Under such an assumption, estimation could proceed by estimating separate hazard models for each of *n* possible outcomes. Gilbert (1992) introduced a competing risks specification and separate estimation for three transitions in a study of automobile ownership duration. Separately estimating competing risks hazards inherently assumes independence among risks. This is frequently done (e.g., Katz, 1986 and Gilbert, 1992) but may not always be appropriate because it ignores potentially important interdependence among risks. Treating competing risks independently is analogous to assuming recursivity in more traditional simultaneous equations problems, which can be solved using three-stage-least squares and similar methods (Hensher and Mannering 1994).

Some researchers also regard the various spells as being analysed as independent events, and apply the methods developed to handle single spells. This is problematic if the populations are heterogenous which would result in a mixing that may lead to a time dependency and incorrect inferences. Since transport applications are characterised by high levels of interdependency between variables, the homogeneity assumption is quite improbable. Incorporating observed and unobserved heterogeneity is necessary, or at least should be tested. Segmentation by socioeconomic characteristics is partially useful - it is however unable to handle the sources of unobserved heterogeneity (and its probable correlation with duration dependence). The importance of introducing time varying covariates and unobserved heterogeneity into a proportional hazards (PH) model is appreciated when it is understood that the PH model, in the presence of time invariant covariates, assumes that the ratio of the hazard for any two sampled members of a population should be constant throughout the observation (i.e. it is independent of time).

Accounting for interdependence among competing risks is not an easy task, but has been undertaken by Diamond and Hausman (1984), Han and Hausman (1990), Sueyoshi (1992) and Meyer (1986,1990). Diamond and Hausman develop a model with strict parametric assumptions on the nature of interdependence. Han and Hausman extend this work by providing a flexible parametric form of interdependence but with time constant covariates, and Meyer (1986) and Sueyoshi (1992) extended the Han-Hausman model to the time varying covariates case. The approach allows one to statistically test whether the more common assumption of independence among competing risks is valid. Meyer (1986) combined non parametric distributions for both components of the hazard function.

Two important issues in the study of event histories are (i) ways of capturing the unobserved heterogeneity in the sampled population (not investigated by Gilbert (1992)) and (ii) the dependency of duration and states over time. These important phenomenon accommodate elements of the dynamics of event histories which influence the nature of transitions. They can be introduced in single spell single state models as well as the more complex multistate multispell models. To introduce these ideas, it is useful to define the information requirements of an event history, and then introduce the essential formulae required to parameterise a competing risks multispell duration model as extensions of equations (5) - (7). Hensher and Mannering (1994) have reviewed the broader literature on duration modelling and applications in transportation, but limited the discussion to single state and single spell methods.

An event history of a sampled household over some observed time period requires information on (i) the initial state (ii) the number of spells in the observation period (iii) the points in time at which some state transition has occurred or a specific event has taken place (iv) state occupancies corresponding to the above points in time (v) an indicator that identifies whether a particular spell is censored and (vi) the set of covariates, measured at the beginning of each spell. Covariates take on three forms: time invariant (e.g. sex), time dependent (e.g. age), and time varying (e.g. lifecycle stage).

If in estimating the hazard rate one aggregates the unobserved differences across the sampled population, an apparent duration dependency occurs. This is potentially spurious duration dependence. At the level of the hazard rate to be analysed it is no longer possible to differentiate whether the hazard rate falls with increasing duration for each household or if this is simply a methodological artefact due to neglected differences between households. While some, hopefully much, of the differences can be accounted for by a set of observed time invariant and/or time varying covariates, there is likely to remain a potentially significant source of unobserved heterogeneity which needs special treatment.

Heterogeneity in general is handled by a mixing distribution over separate (but jointly estimated) hazard functions. A popular way of incorporating heterogeneity is as a random multiplicative factor that shifts the baseline hazard:

$$h_o(t) = \exp(\theta_o \lambda_b(t)) \exp(z_o(t)\beta) \tag{8}$$

where θ_o is a random variable associated with initial state *o* with a distribution defined by the analyst, representing the distribution of the unobserved heterogeneity within the population of sampled households. The random variable must be limited to positive values (given that the hazard rate is not negative). If we set $E(\theta) = 1$, then on average one obtains $\lambda_b(t)$. Parametric specifications have been investigated, especially the gamma, normal, and logistic mixing distributions. Heckman and Singer (1984) proposed a non parametric finite mixture model to accommodate the highly sensitive nature of the parameter estimates associated with the covariates to alternative distributional assumptions. A non parametric specification for the heterogeneity profile (or *finite* mixture model) is defined by a set of support values (typically up to 10), which are estimated jointly with the probability mass for each point. Unlike the parametric specification of θ_o, a fraction of the population can have a zero hazard rate. There is considerable debate in the literature as to whether the baseline hazard or the mixture distribution should be non parametric. Trussell and Richards (1985) for example, suggest that a non parametric baseline and a parametric mixture distribution are equally plausible. This topic is ripe for extensive empirical inquiry. We investigate the implications of parametric and non-parametric specifications of unobserved heterogeneity in our empirical application. We impose the assumption that there is no omitted variable bias, due to the correlation of any observed covariates and unobserved heterogeneity. The possibility of dependency could be tested by specifying θ_o as a function of the covariates. This greatly complicates the model, including the possibility of identification problems.

The hazard model for a competing risk model can be defined as:

$$h(t_{od}|\mathbf{z},\theta) = \exp\{\delta_{odb} + z_o(t_{od} + \tau)\beta_{od} + \sum \delta_{odk} f_k(t_{od}) + c_{od}\theta\} \tag{9}$$

and

$$f_k(t) = \sum_{k=1}^{K} (t_{od}^{\lambda_k} - 1)/\lambda_k; \quad \lambda_k = \mu_{odk} \tag{10}$$

where δ_{odb} is the baseline hazard for a multi-state model, $z_o(t_{od})$ defines time-varying covariates, $z_o\tau$ defines time dependent covariates, and $f_k(t_{od})$ is defined by equation (10) as a Box-Cox transformation over time to capture general duration dependence. Setting K = 1 and hence $\lambda_1 = 0$ gives a Weibull distribution; setting K= 1 and $\lambda_1 = 1$ gives a Gompertz distribution. Other functional forms are possible. For example, setting K = 2 with $\lambda_1 = 1$ and $\lambda_2 = 2$ produces a quadratic duration dependence. Lillard (1993) chose a piecewise linear spline to represent the dependence of hazards on calendar time. $c_{od}\theta$ is a weighted unobserved heterogeneity index, where θ is common across all transitions o to d, and the weight, c_{od} conditions the unobservable scalar to have a differentiating role in different transitions or different spells. Equation (9) is a very general specification of a hazard function allowing for time varying covariates, unobserved heterogeneity and duration dependence. Setting $\beta_{od} = c_{od} = f_{k\,(t)} = 0$ gives an exponential form for the hazard function. Parametric or non-parametric assumptions can be imposed on θ as discussed above.

Equation (9) is the kernel of the specification of a multistate multispell model with allowance for time varying covariates, unobserved heterogeneity and duration dependence. The challenge now is to estimate a number of hazard functions under the most interesting alternative specifications. In the context of the household's timing and duration of automobile transactions, four empirical model specifications are investigated:

M1: parametric baseline hazard, time varying covariates, no unobserved heterogeneity, duration dependence
M2: parametric baseline hazard, time varying covariates, unobserved heterogeneity, duration dependence
M3: parametric baseline hazard, no time varying covariates, no unobserved heterogeneity, duration dependence
M4: parametric baseline hazard, no time varying covariates, unobserved heterogeneity, duration dependence

In models M2 and M4 we investigate one parametric distribution - log normal - and a non-parametric finite mixture model for unobserved heterogeneity. Duration dependence is evaluated under Weibull, and Gompertz distributions.

3. AN EMPIRICAL STUDY OF AUTOMOBILE TRANSACTIONS

A sample of 200 households from the Sydney automobile panel (Hensher et al. 1992) who provided complete information over a 12 year period on a limited number of socioeconomic and vehicle characteristics (Table 1) were used in the empirical application. The data for the years 1981-1984 were obtained from an annual reinterview; the other data (1980-1974) were collected retrospectively at the conclusion of the panel. Given the problems associated with retrospective data, the number of items of data obtained were somewhat limited, but adequate for the current purpose. The sample sizes for each transition are: 1 to 2 = 197, 1 to 3 = 137, 2 to 1 = 212, 3 to 1 = 140, 1 to 1 = 163, 2 to 2 = 23, 3 to 3 = 14, 2 to 3 = 0, and 3 to 2 = 0. This gives 886 spells.

The importance of understanding the timing and duration of automobile transactions is well documented (eg. Kitamura 1987, 1989, Smith et al. 1991). Despite this recognition, the empirical efforts are few. The only substantive study is by Gilbert (1992), although the interest in automobile transactions modelling is growing. Gilbert treated each transition as independent events and ignored unobserved heterogeneity. The current study is the only known application of competing risk multispell models in transportation in which the transitions are estimated jointly with allowanc for unobserved heterogeneity. The purpose is to identify the influences on the probability that a sampled household will undertake a particular type of transaction over the period 1974-85 given the observation of one of three states in each time interval. The three states are no change, replace a used vehicle with a used vehicle, and replace a used vehicle with a new vehicle. Out of 2400 observations across 200 households and 12 years, we have 2011 (83.8%) states of no change, 235 (9.8%) replacements with a used vehicle and 154 (6.4%) replacements with a new vehicle. In this paper we limit the empirical assessment to joint estimation of transitions 1 to 2 and 1 to 3. The average duration of the transition from no change to replace with a used vehicle is 3.90 years; the equivalent mean for a replacement with a new vehicle is 4.41 years.

No.	Acronym	Definition	Mean (sd)
1	END	End of case identifier (1,0)	
2	YR	Year (74,75,76,....,85)	
3	STATE	State (1 = no change, 2 = replace used with used vehicle, 3 = replaced used with new vehicle)	
4	HSIZE	Household size (	2.96 (1.44)
5	NHINC	Number of Income earners in household	1.67 (0.64)
6	LIFA	lifecycle A (1,0) young adults (<35), no children	0.053
7	LIFBCD	lifecycle BCD (1,0) two heads, children up to 12 years old	0.196
8	LIFEF	lifecycle EF (1,0) one or two heads, children over 16 years	0.191
9	LIFG	lifecycle G (1,0) older adults, no children	0.228
10	LIFH	lifecycle H (1,0) retired persons over 65 years old	0.226
11	LIFIJ	lifecycle IJ (1,0) single head,	0.107
12	RGHH	1 or more vehs. are private registered (1,0)	0.705
13	REGHS	1 or more vehs are household business registered (1,0)	0.171
14	REGOT	1 or more vehs are other company registered (1,0)	0.127
15	LOCAL	Prime county of manufacture (1 = local, 0 = other)	0.491
		12 years of data for a sample of Sydney (Australia) households, 2400 lines of data or 200 households	

Table 1. The Data Set Used in Model Estimation

There are 5 time varying covariates - household size (HSIZE), number of income earners (NHINC), household stage in the lifecycle (LIF..), number of vehicles in each registration category (REG..) and the prime country of vehicle manufacture (LOCAL). Within the limits of the data a number of broad issues are worthy of investigation. In particular we want to evaluate the role that changing household life cycle and vehicle registration status plays in the households automobile replacement decision. To what extent are households loyal to the used car market or are willing to trade up to new vehicles? Automobile manufacturers are particularly interested in this question as might be proponents of alternative fuelled vehicles in the early formative years. Since there is almost certainly likely to be some important missing covariates, allowance for unobserved heterogeneity will be important to the results.

The set of models estimated under different assumptions on the form of duration dependence and unobserved heterogeneity for a given set of significant time varying covariates are summarised in Table 3. The set of possible model forms is extensive. We have limited Table 3 to a sufficiently broad range of situations to illustrate the diversity of results. The Weibull and Gompertz distributions provide a good array of alternative interpretations of behavioural response over time (see Hensher and Mannering 1994 for further details) for duration dependence. Under a Box-Cox specification of duration dependence (equation 10), we set K = 1 and lambda equal to 0 and 1 respectively for Weibull and Gompertz distributions.

The Weibull distribution is a generalised form of the exponential distribution. The Weibull distribution imposes the monotonicity restriction on the hazard. We are able to identify whether loyalty to the used car market is time-dependent or time-independent. The Gompertz distribution, derived from the extreme-value distribution, is truncated at zero so that no negative values are possible. Unobserved heterogeneity is evaluated as a parametric lognormal distribution and as a non-parametric mixture specification. We have assumed 10 intervals on each side of the mean to approximate the lognormal distribution.Since the distribution is asymmetric, the intervals will be of different lengths on each side of the mean. A non-parametric cumulative density function with 3 support points on the unit interval is specified with all of the support points and cumulative probabilities fixed. Allowing free estimation of a range of support points, except the first and last points and the last cumulative probability, gave spurious results. Further investigation is warranted.

3.1 Discussion of Illustrative Results

The hazard of replacing a vehicle with a used vehicle (transition 1 to 2) or with a new vehicle (transition 1 to 3) varies quite noticeably between the transition types and the distributional assumptions on duration dependence and unobserved heterogeneity. Beginning with no unobserved heterogeneity, the shape parameter (gamma) for duration dependence for both distributions is significantly positive in all models across both transitions suggesting that for both distributions the hazard is an increasing function of time. When we control for unobserved heterogeneity the shape parameter has a stronger influence on the hazard, increasing the expected time in a state, ceteris paribus.

Adding in a set of time-varying covariates to remove the role of life cycle stage, number of income-earning household members, and the registration status of the household vehicles (a proxy for financial obligation in vehicle transactions) has very little impact on the scale (i.e. constant) and shape parameters. The exception appears to be for the model with a Weibull duration dependence and non-parametric unobserved heterogeneity. Here we find that the scale parameter changes quite substantially for both transitions suggesting some influence of time varying covariates. A closer examination however highlights the change in the middle support point which was statistically significant in the full model and insignificant in the absence of the covariates. The reason for this is unclear. One might postulate that the non-parametric distribution in the presence of the covariates is ‘about right’ but completely unsuitable when it has to carry more information.

The only three covariates approaching acceptable statistical significance are REGHS (household has at least one household-business registered vehicle) in transition 1 to 2 , and LIFBCD (households in lifecycle stage of two heads and children up to 12 years old) and LIFG (households with older adults and no children) in transition 1 to 3 for Gompertz duration dependence and non-parametric unobserved heterogeneity. The negative sign on REGHS suggests that the hazard of replacement with a used vehicle decreases, ceteris paribus, where households have access to a household-business registered vehicle relative to a privately registered vehicle. The life cycle effects are both positive implying that a household in either of these life cycle stages, ceteris paribus, has a higher hazard of replacement with a new vehicle.

A useful way of comparing the alternative specifications is to tabulate the hazard as a function of time. Given the statistical insignificance of the covariates we limit this to the models containing the scale, duration shape and unobserved heterogeneity parameters (Table 2). The predicted hazards in parenthesis relate to parametric unobserved heterogeneity. The Weibull and Gompertz specifications are monotonically increasing in duration implying that the longer a household goes without exiting a duration, the more likely it is to exit soon. The effect is stronger for transition 1 to 3 than transition 1 to 2. The turnover is greater for used vehicles than new vehicles. For transition 1 to 2, the hazard is higher for the Weibull distribution for 2 to 6 years with the Gompertz producing a greater hazard for 7 to 10 years. For transition 1 to 3 the Gompetz has the higher hazard up to 2 years and after 7 years with the Weibull higher in the middle time durations. When allowance is made for unobserved heterogeneity we find some re-ordering of relativities and some significant adjustments in the hazard for transition 1 to 3: allowance for unobserved heterogeneity reduces the hazard with the gap increasing as duration increases. The difference for transition 1 to 2 is not noticeable at all. This leads one to conclude that failure to control for unobserved heterogeneity tends to lead to an over-estimate of the hazard for transitions involving replacement of a vehicle with a used vehicle, but its has no effect in the new car market.

Time (years)	DD=Weibull 1 to 2	DD=Weibull 1 to 3	DD=Gompertz 1 to 2	DD=Gompertz 1 to 3
1	0.038 (0.037)	0.012 (0.009)	0.052 (0.051)	0.028 (0.023)
2	0.077 (0.076)	0.036 (0.029)	0.070 (0.079)	0.041 (0.039)
3	0.116 (0.115)	0.071 (0.056)	0.094 (0.102)	0.059 (0.052)
4	0.156 (0.155)	0.114 (0.090)	0.126 (0.129)	0.087 (0.072)
5	0.196 (0.195)	0.165 (0.130)	0.170 (0.173)	0.127 (0.113)
6	0.237 (0.235)	0.223 (0.175)	0.229 (0.230)	0.185 (0.162)
7	0.277 (0.275)	0.287 (0.226)	0.309 (0.311)	0.270 (0.235)
8	0.318 (0.315)	0.358 (0.282)	0.416 (0.417)	0.395 (0.346)
9	0.358 (0.355)	0.435 (0.343)	0.560 (0.558)	0.577 (0.523)
10	0.399 (0.396)	0.518 (0.408)	0.755 (0.753)	0.843 (0.721)

Table 2. Estimated Hazard Functions

	Variables	DD=Weibull 1 to 2	DD=Weibull 1 to 3	DD=Gompertz 1 to 2	DD=Gompertz 1 to 3
UH=0	constant	-3.139 (-9.33)	-4.301 (-10.4)	-2.820 (-8.65)	-3.424 (-9.71)
	gamma	1.069 (9.22)	1.723 (8.2)	0.314 (9.50)	0.399 (7.94)
	lambda	0.00	0.00	1.00	1.00
	nhinc	-0.117 (-.77)	-0.109 (-.71)	-0.105 (-.72)	-0.082 (-.53)
	lifa	0.624 (1.46)	-0.021 (-.04)	0.598 (1.31)	-0.083 (-.13)
	lifbcd	0.125 (0.45)	0.128 (0.46)	0.083 (0.30)	0.098 (0.35)
	lifef	-0.045 (-.17)	-0.373 (-1.3)	-0.085 (-.33)	-0.038 (-1.31)
	lifg	0.112 (0.44)	0.215 (0.75)	0.095 (0.37)	0.191 (0.67)
	lifij	0.101 (0.33)	-0.497 (-1.37)	0.103 (0.36)	-0.483 (-1.34)
	regot	-0.195 (-.67)	-0.011 (-.03)	-0.126 (-.42)	0.052 (0.16)
	reghs	-0.392 (-1.7)	-0.129 (-.56)	-0.414 (-1.80)	-0.153 (-.64)
	LL (0)	-1029.30		-1057.74	
	LL (C)	-1021.46		-1052.30	
UH=0	constant	-3.278 (-18.5)	-4.462 (-13.95)	-2.963 (19.09)	-3.582 (-16.04)
	gamma	1.025 (9.49)	1.652 (8.52)	0.298 (9.54)	0.379 (8.27)
	lambda	0.00	0.00	1.00	1.00
	LL (0)	-1069.32		-1061.8	
	LL (C)	-1030.86		-1061.3	
UH =lognormal	constant	-3.03 (-7.31)	-4.086 (-8.47)	-2.878 (-2.46)	-3.535 (-2.51)
	gamma	1.112 (8.82)	1.947 (8.10)	0.325 (7.24)	0.403 (5.40)
	lambda	0.00	0.00	1.00	1.00
	nhinc	-0.127 (-.81)	-0.150 (-.85)	-0.117 (-.78)	-0.085 (-.53)
	lifa	0.638 (1.42)	-0.008 (-.01)	0.609 (1.32)	-0.060 (-.09)
	lifbcd	0.156 (0.55)	0.214 (0.70)	0.085 (0.31)	0.103 (0.36)
	lifef	-0.037 (-.14)	-0.370 (-1.15)	-0.094 (-.36)	-0.377 (-1.26)
	lifg	0.128 (0.50)	0.251 (0.82)	0.093 (0.36)	0.191 (0.65)
	lifij	0.090 (0.29)	-0.547 (-1.44)	0.122 (0.43)	-0.473 (-1.29)
	regot	-0.223 (-.75)	-0.085 (-.25)	-0.131 (-.44)	0.045 (0.13)
	reghs	-0.375 (-1.49)	-0.052 (-.20)	-0.427 (-1.78)	-0.157 (-.66)
	factor loading	-0.103 (-.61)	-0.341 (-1.24)	0.035 (0.05)	0.070 (.09)
	LL (0)	-2132.55		-1047.72	
	LL (C)	-1016.11		-1047.45	
UH= lognormal	Constant	-3.187 (-10.4)	-4.293 (-9.4)	-2.806 (-16.3)	-3.357 (-12.2)
	gamma	1.057 (9.36)	1.778 (8.03)	0.403 (11.1)	0.649 (10.7)
	lambda	0.00	0.00	1.00	1.00
	factor loading	-0.084 (-.45)	-0.239 (-.78)	-0.278 (-3.09)	-0.785 (-3.91)
	LL (0)	-1218.63		-1060.13	

	LL (C)	-1025.85		-1046.30	

Table 3 continued

UH=non-parametric	constant	-2.594 (-3.34)	-2.484 (-1.84)	-1.796 (-3.50)	-0.180 (-.27)
	gamma	1.144 (8.95)	2.104 (8.56)	0.422 (9.61)	0.768 (12.3)
	lambda	0.00	0.00	1.00	1.00
	nhinc	-0.135 (-.83)	-0.184 (-1.00)	-0.131 (-.81)	-0.193 (-.99)
	lifa	0.651 (1.42)	0.0126 (0.02)	0.707 (1.52)	0.148 (0.24)
	lifbcd	0.183 (0.63)	0.310 (0.95)	0.238 (0.84)	0.544 (1.70)
	lifef	-0.023 (-.09)	-0.334 (-1.02)	-0.017 (-.06)	-0.200 (-.65)
	lifg	0.139 (0.54)	0.293 (0.94)	0.203 (0.76)	0.519 (1.72)
	lifij	0.092 (0.29)	-0.553 (-1.45)	0.095 (0.32)	-0.459 (-1.22)
	regot	-0.244 (-.82)	-0.146 (-.43)	-0.278 (-.94)	-0.351 (-1.127)
	reghs	-0.381 (-1.44)	-0.033 (-.12)	-0.447 (-1.77)	-0.176 (0.67)
	factor loading	-1.053 (-.79)	-4.014 (-1.45)	-2.152 (-3.17)	-7.614 (-6.20)
	support point	0.841 (4.58)	0.841 (4.58)	0.805 (17.99)	0.805 (17.99)
	LL (0)	-1021.46		-1052.30	
	LL (C)	-1020.33		-1033.38	
UH= non-parametric	Constant	-3.276 (-.00)	-4.457 (-.016)	-1.923 (-5.21)	-0.681 (-1.20)
	gamma	1.025 (8.85)	1.652 (5.65)	0.402 (10.3)	0.682 (12.5)
	lambda	0.00	0.00	1.00	1.00
	factor loading	-0.002 (-.00)	-0.005 (-.01)	-2.187 (-3.17)	-6.705 (-6.10)
	support point	0.004 (0.01)	0.004 (0.01)	0.819 (17.4)	0.819 (17.4)
	LL (0)	-6724.37		-1065.36	
	LL (C)	-1030.87		-1047.64	

Table 3. Illustrative Model Results for Alternative Specifications

4. CONCLUSIONS

Event history data embedded in some panel or activity diary data sets in transportation offer an opportunity to investigate the underlying structure of duration that a household is in a particular state and the timing of a change into another state. The literature on multistate multispell modelling in continuous time offers a future prospect for improving our understanding of ‘when’ changes are likely to occur. The consequences for improved forecasting of change into the future is clear. Existing methods of modelling travel behaviour, including recent dynamic discrete choice models, are limited in the advice they give on the timing of change. Knowing if a change will occur is handicapped if we lack a procedure for identifying when it will occur. There is a lot more research required to increase our empirical knowledge of the implications of alternative assumptions on how duration dependence, unobserved heterogeneity and baseline hazards are specified. The illustrative empirical study highlights the role of these various dimensions in establishing a capability for predicting the timing of change. This paper is a contribution to this effort in transportation. The greatest challenge however will continue to be the establishment of sufficiently rich data sources capable of assisting the transport analyst in the search for improved methods of forecasting the duration and timing of change, and the establishment of efficient software capable of modelling the myriad of competing risks and multiple spell event histories in transportation.

ACKNOWLEDGMENTS

The comments, advice and assistance of Phillip Colla, Fred Mannering, David Brownstone, Bill Greene and Dan Steinberg are greatly appreciated. This paper was written while a Visiting Professor at the Institute of Transportation Studies, University of California at Irvine.

REFERENCES

Axhausen, K.W. (1992) **British Travel Behaviour Panels**, Department of Civil Engineering, Imperial College, London.

Diamond, P. and Hausman, J. (1984) The retirement and unemployment behavior of older men. In H. Aaron and G. Burtless (eds), **Retirement and Economic Behavior**, Brookings Institute, Washington, DC.

DuWors, R.E. and Haines, G.H. (1990) Event history analysis measures of brand loyalty, **Journal of Marketing Research**, XXVII, 485-493.

Econometric Software Inc (1993) **LIMDEP 6.0**, Econometric Software, New York and Sydney.

Gilbert, C.C.M. (1992) A duration model of automobile ownership, **Transportation Research-B**, 26B(2): 97-114.

Golob, T.F. and Golob, J.M. (1989) **Practical considerations in the development of a transit users panel**, International Conference on Dynamic Travel Behaviour Analysis, Kyoto.

Hamed, M.M. and Mannering, F.L. (1992) A note on commuters' activity duration and the efficiency of proportional hazards models, **Department of Civil Engineering, University of Washington**, Seattle (mimeo).

Han, A. and Hausman, J. (1990) Flexible parametric estimation of duration and competing risk models. **Journal of Applied Econometrics**, 5, 1-28.

Hannan, M.T. and Tuma, N.B.(1979) Methods of temporal analysis, **Ann. Rev.Sociol**, 5, 303-328.

Heckman, J. and Singer, B. (1984) A method for minimising the impact of distributional assumptions in econometric models for duration data, **Econometrica**, 52, 271-320.

Hensher, D.A. and Raimond, T. (1992) The Timing of Change: Discrete and Continuous Time Panels in Transportation, Presented at the **First U.S.A. Invitational Conference on**

Panels in Transport Planning, UCLA Conference Center, Lake Arrowhead, California, October 25-27 (to appear in the Proceedings by Kluwer Publishers, Amsterdam).

Hensher, D.A. and Mannering, F.L. (1994) Hazard-Based Duration Models and Their Application to Transportation Analysis, **Transport Reviews**, 14 (1), 63-82.

Hensher, D.A., Smith, N.C., Milthorpe, F.M. and Barnard, P.O. (1992) **Dimensions of Automobile Demand: A Longitudinal Study of Automobile Ownership and Use**, Amsterdam, North Holland.

Kalbfleisch, J. and Prentice, R. (1980) **The Statistical Analysis of Failure Time Data**, John Wiley & Sons, New York.

Katz, L. (1986) Layoffs, recall and the duration of unemployment. **Working paper No. 1825,** National Bureau of Economic Research, Cambridge, MA.

Kitamura, R. (1987) A panel analysis of household car ownership and mobility, **Proceeding of Japan Society of Civil Engineers**, 338/IV-7 (Infrastructure Planning and Management), 13-37.

Kitamura, R. (1989) A causal analysis of car ownership and transit use, **Transportation**, 16(2): 155-173.

Lancaster, T. (1979) Econometric methods for the duration of unemployment, **Econometrica**, 47, 939-956.

Lillard, L.A. (1993) Simultaneous equations for hazards: marriage duration and fertility timing, **Journal of Econometrics**, 56, 189-217.

Meyer, B. (1986) Semi parametric estimation of hazard models, unpublished paper, **Northwestern University**, Evanston.

Meyer, B. (1990) Unemployment insurance and the distribution of unemployment, **Econometrica**, 58, 757-82.

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

Paselk, T.A. and Mannering, F.L. (1992) Use of duration models for predicting vehicular delay at U.S./Canadian border crossings, paper for presentation at the **1993 Annual Transportation Research Board Meeting,** Washington D.C.

Prentice, R.L. and Gloeckler, L.A. (1978)Regression analysis of grouped survival data with applications to breast cancer data, **Biometrics**, 34, 57-67.

Raimond, T. and Hensher, D.A. (1992) Panel Surveys and Other Longitudinal Techniques: An Annotated Bibliographic Review, Working Paper ITS-WP-92-19 Proceedings of the **First U.S.A. Invitational Conference on Panels in Transport Planning**, UCLA Conference Center, Lake Arrowhead, California, October 25-27 (Kluwer Publishers, Amsterdam, in press).

Smith, N.C., Hensher, D.A. and Wrigley, N. (1991) Modelling Discrete Choice Outcome Sequences with Panel Data: An Application to Automobile Transactions, **International Journal of Transport Economics**, XVIII (2), June, 123-150.

Sueyoshi, G. (1992) Semi parametric proportional hazards estimation of competing risks models with time-varying covariates, **Journal of Econometrics**, 51, 25-58.

Trussell, J. and Richards, T. (1985) Correcting for unobserved heterogeneity in hazard models: an application of the Heckman-Singer model to demographic data, **Sociological Methodology**, 242-276.

van Wissen, L.J.G. and Meurs, H.J. (1989) The Dutch Mobility Panel: experiences and evaluation, **Transportation**, 16(2): 99-119.