Time taken for residents to adopt a new public transport service: examining heterogeneity through duration modelling

KIRON CHATTERJEE[1*] & KANG-RAE MA[2]
[1] Centre for Transport & Society, Faculty of the Built Environment, University of the West of England, Bristol, BS16 1QY, UK
E-mail: Kiron.Chatterjee@uwe.ac.uk
Tel: 00 44 117 32 82032
Fax: 00 44117 32 83002

[2] Department of Urban and Regional Planning, Chung-Ang University, Anseong-si, Gyeonggi-do, Korea
E-mail: kma@cau.ac.kr
Tel: 00 82 31 670 4832
Fax: 00 82 31 675 1381

*Author for correspondence

Contributors:
Kiron Chatterjee has been a Senior Lecturer at the University of the West of England, Bristol, since 2003 and previously was at the University of Southampton. Currently, a main focus of his research is on longitudinal analysis of travel behaviour to improve policy analysis.

Kang-Rae Ma received a PhD in Planning from University College London. He worked at the University of the West of England and the Korea Transport Institute before he joined Chung-Ang University as an Assistant Professor. His research interests include modelling of travel behaviour and urban excess commuting.

Key words: panel data, travel behaviour, dynamics, response lag, duration modelling.

Abstract: When a new public transport service is introduced it would be valuable for public authorities, financing organisations and transport operators to know how long it will take for people to start to use the service and what factors influence this. This paper presents results from research analysing the time taken for residents living close to a new guided bus service to start to use (or adopt) the service. Data was obtained from a sample of residents on whether they used the new service and the number of weeks after the service was introduced before they first used it. Duration modelling has been used to analyse how the likelihood of starting to use the new service changes over time (after the introduction of the service) and to examine what factors influence this. It is found that residents who have not used the new service are increasingly unlikely to use it as time passes. Those residents gaining greater accessibility benefits from the new service are found to be quicker to use the service, although the size of this effect is modest compared to that of other between-resident differences. Allowance for the possibility that there existed a proportion of the sample that would never use the new service was tested using a split population model (SPD) model. The SPD model indicates that 36% of residents will never use the new service and is informative in differentiating factors that influence whether Route 20 is used and when it is used.

1. Introduction

When new options are introduced into the transport market or existing options are modified, travel demand responses are not instantaneous, but evolve over time. Those concerned with the provision of transport services (e.g. public authorities, financing organisations, transport operators) will be interested in how many new users they can attract and how quickly they can be attracted.

Conventional methods of travel demand analysis (based on cross-sectional travel data and on equilibrium principles) are static in nature and are not able to forecast the evolution of demand for a new transport service. They assume that travel demand will attain a new level after the service is introduced, but do not indicate any time-scale for when this level of demand will be reached. It is the dynamic demand profile that will determine the consequences of a new transport service for public welfare (user benefits, societal costs) and business viability (revenue streams).

Douglas (2003) has carried out an analysis of the patronage growth for 13 new or upgraded rail schemes from around the world and estimated an average 'ramp-up' factor of 79% for the first year of operation, 95% for the second year of operation and steady state patronage after three years. However, there was considerable variation in growth across the schemes. To address the shortcomings of static forecasts from conventional transport models, in May 2002 the UK Department for Transport (DfT) issued general advice for major public transport schemes to assume that 80% of full patronage build-up is attained by the end of year 1, 90% by the end of year 2, 95% by the end of year 3 and 100% by the end of year 4. In April 2003, a revised version of the advice document omitted these default values (DfT 2003; Appendix B, para. B36). DfT stated that they required build-up values to be used which reflect the particular scheme and suggested to scheme promoters that they refer to the experience of similar schemes introduced elsewhere or seek advice from DfT. The change in the position taken by DfT highlights the lack of current knowledge on the timing of traveller responses to a change in the travel market.

It is apparent that some individuals will respond quickly, or immediately, in using or *adopting* a new service, while others take longer before adopting the service or do not adopt it at all. Objective reasons for faster adoption are that the new service saves time or money for a journey that an individual makes, or that the new service provides good access to a destination that an individual has not been able to reach previously. Subjective reasons for faster adoption are that an individual is aware of the new service in advance and makes plans to use it, or that an individual is prompted to deliberate about their travel choices by an external event (e.g. unexpected delays using a current mode).

In order to gain insights on the timing of responses to new public transport services, a panel survey was organised to coincide with the introduction of a new guided bus service in Crawley, Southern England. Data was obtained from a sample of residents on whether they used the new service and the number of weeks after the service was introduced before they first used it. Duration models are used in this paper to analyse the *elapsed time* until residents adopted the new bus service. In particular, duration models are used to investigate how the probability of starting to use the new bus service varies over time (after its introduction) and how this is affected by between-resident differences in personal, household and travel characteristics. A previous paper by the authors introduced the panel survey and included some initial duration modelling results from the data (Chatterjee and Ma 2007), but this paper considers different modelling specifications that can be used to examine heterogeneity in the responses of residents.

In the next section of the paper the context for the panel survey and the survey itself are briefly described. In section 3, an overview is provided of the duration modelling approach before the analysis sample and duration data set are described in section 4. The model specifications tested are explained in section 5 with results presented and discussed in section 6. Finally, the results of the study are summarised and conclusions are drawn in section 7.

2. Panel survey

Fastway Route 20

The Fastway bus system began operating in the Crawley and Gatwick Airport area in the county of West Sussex, Southern England, in September 2003 (Fastway 2008). It is intended to be a modern, high quality public transport system providing a frequent, reliable service and offering a real alternative to the car. The Fastway buses travel in dedicated lanes and guideways along significant parts of their routes and also benefit from barrier controlled bus gates and priority at signal controlled junctions. Real-time information is provided at bus stops and on the internet and the buses are a modern fleet of high specification vehicles with low floor access, comfortable and modern interiors and low-noise and low-emission engines.

The Fastway system supplements existing bus services within the area and is designed to provide more direct public transport services than otherwise available connecting residential areas with key employment sites such as Crawley Town Centre and Gatwick Airport. The first Fastway service (Route 10) experienced steady growth in passengers from 4,000 passengers per day in September 2003 up to 6,000 in May 2005 and the second service (Route 20) was introduced in August 2005. The route maps are shown in Figure 1. It is the Route 20 service that provides the case study for this paper.

[Figure 1 here]

Crawley Panel Survey

The survey aimed to obtain information on travel behaviour for a sample of residents living close to the new service over a period of time before and after the introduction of the service. In particular, the survey sought to track usage of the new Route 20 service and to identify as accurately and precisely as possible when residents first used Route 20. A classic panel survey has been conducted involving the same respondents being surveyed at four different time points. Event history data recording behaviour in continuous time was not a feasible option.

Douglas (2003) identified an average of 79% 'ramp-up' one year after a rail scheme introduction or upgrade. This suggests that most but not all responses take place within a year. For a new bus service the time-scale of responses is likely to be shorter than for a new rail service, as it will be used for local journeys and is likely to be more readily known to potential users. The Crawley panel survey involved four waves with wave one taking place one month before the introduction of the Route 20 service, wave two taking place one month after the introduction of the service, wave three taking place three months after the introduction of the service and wave four taking place six months after the introduction of the service. The overall length of the panel survey was chosen to span a sufficient period for behavioural responses to take place and start to diminish. The intervals between waves were chosen to be sufficiently short in duration in order to identify approximately when a change in behaviour occurred.

The target population for the panel survey was residents living close to the route of the new Route 20 service and not living close to the route of the earlier introduced Route 10 service. The locations for the target population in Broadfield (south) and Three Bridges (east) are shown by the shaded areas in Figure 1. These residents gained from Route 20 a significantly faster public transport connection to key destinations in the area. Characteristics of the two targeted areas are presented in Table 1.

[Table 1 here]

The electoral register was used to identify residents in the target areas with the register providing names and addresses of approximately 2,500 residents. The panel survey used self-administered postal questionnaires as the survey instrument. The first wave of the postal survey (in August 2005) achieved a 22% response rate (554 responses) which is typical of experience with self-completion postal questionnaires. 361 respondents said they were willing to participate further in study. These were sent the second questionnaire and 220 responses were received (in October 2005). To maximise subsequent participation a £20 incentive was offered to those participating in the final two waves. 254 responses were received for wave 3 (in December 2005) and 247 responses were received for wave 4 (in March 2006). No attempt was made to refresh the sample during the course of the study, due to there being no further source of participants. Possible sample biases arising from initial non-response and attrition are discussed in section 3.

The structure and design of the questionnaire was similar in each wave to ensure as far as possible that responses were directly comparable. Respondents were asked to provide various information including personal and household information, frequency of use of different transport modes and specific information about use of Route 20. Residents were asked in waves 2, 3 and 4 to indicate if they had used the service and in which preceding week they had first used the service. It is recognised that respondents may not easily recollect this, but accuracy to within a month is likely given the two month intervals between survey occasions.

3. The duration modelling approach

Of interest in this study is the elapsed time, after Route 20 is introduced, until residents started to use the service. Duration data has been obtained, recording the length of time (to the nearest week) survey respondents spent without using Route 20, or otherwise recording that they had not used Route 20 by the end of the survey period.

Hazard-based duration models have been used to analyse the data. These models examine the conditional probability of a duration ending at time t, given that the duration has continued until time t. This is known as the hazard probability. Hazard-based duration models allow survival functions to be obtained which give the probabilities of durations enduring (or surviving) until time t. Survival functions therefore indicate the delayed uptake in use of the service. As well as accounting for duration dependence, duration models can also account for the effect of covariates on hazard probabilities. In this study, duration models are used to examine how the probability of starting to use the new bus service changes over time (after its introduction) and how this is affected by characteristics of the residents and the change in accessibility that they experience.

The first research hypothesis to be tested in the analysis is that the **hazard probability of using Route 20 declines over the measurement period.** It is considered that residents who have not used the service become increasingly unlikely to use the service as time goes on, perhaps due to the service not suiting their travel needs or them having negative attitudes towards buses. The second hypothesis to be tested is that the **hazard probability of using Route 20 increases for those residents gaining greater accessibility benefits from Route 20.** It is considered that residents who gain greater accessibility benefits are more likely to start using the service sooner.

The statistical foundations of duration modelling (alternatively known as survival modelling or event history modelling) can be found in texts such as Yamaguchi (1991), Box-Steffensmeier and Jones (1997), Le (1997) and Jenkins (2004). Hensher and Mannering (1994), Bhat (2000) and Washington et al. (2003) explain how hazard-based duration models can be applied to transport problems. They note that there have been surprising few applications of duration modelling in the transport field, but that these have been increasing recently. One example where hazard-based duration modelling has been applied to the time to adopt a new transport alternative is reported by Hensher (1997). This concerned the elapsed time until motorists switched to a new toll road. It was found that the longer a motorist had not switched to the toll road, the more likely they were to do so. Two covariates were tested. It was found that the greater the time savings for using the toll road the more likely motorists were to switch to the toll road, and it was found that motorists with company cars were more likely to switch. In this study of public transport adoption a larger ser of covariates will be tested.

Before proceeding to describe the data and analysis, it is noted that the panel survey also provided data (for each respondent at each survey wave) for the weekly frequency of use of

Route 20. From this data, dynamic choice models can be estimated that predict for individuals how frequency of Route 20 use changes over time and what influences this. By recognising persistence of Route 20 use, these models can be used to forecast patronage growth within and beyond the survey period. The argument, though, for analysing duration times in this paper is that it allows specific attention to be devoted to timing of adoption of the new bus service and it thereby enables the full information available on duration times to be used for this purpose. As noted by Peterson (1991), an indicator variable at discrete points cannot capture the range of variability in duration times and leads to estimates that are inefficient (have larger variances).

4. Analysis sample and data set

Analysis sample

The sample used for duration analysis is the residents who continued to participate in the panel survey beyond at least the third wave and thus offered information on duration times for at least three months after Route 20 introduction. This results in an analysis sample of 247 respondents.

It cannot be argued that the analysis sample is representative of the population from which it was drawn for two main reasons. Firstly, the sampling frame (publicly available 'edited' electoral register) used for the survey did not contain a complete list of adults, since households can opt out from having their details available (typically 30% of households opt out). The individuals that opt out may be disproportionately drawn from particular groups of the population. Secondly, a high proportion of contacted individuals did not participate in the survey. At wave 1 (before Route 20 introduction) a 22% response rate was achieved. Of these, about half (247) continued in the panel survey and provided duration times. These participants cannot be expected to be representative of the wider population.

Weighting of the data can be applied to take into account over- and under-representation of sub-groups of the population. Weighting is not straightforward to apply, though, for taking into account of incomplete sampling frames and survey non-response. Recommended methods are based on weighting survey data according to published population distributions by age, sex and geographical area (Cabinet Office 2004). For this analysis no attempt has been made to weight the data in this way. This can be justified as it was not the expressed aim of the panel study to obtain a representative sample of the study area population. Instead of seeking statistical generalisations, the panel survey was intended to generate a sample with sufficient between-person differences in characteristics to enable heterogeneity in timing responses to be explored. It should be acknowledged, though, that heterogeneity in the panel sample is likely to less than in the wider population.

Table 2 compares the characteristics of the analysis sample, the 554 respondents to wave 1 of the panel survey and the population of the town of Crawley. Comparison of the analysis sample to the Crawley population indicates a reasonable match, except that the analysis sample under-represents young adults and individuals in households without a car. Comparison of the analysis sample to the wave 1 sample indicates that the burden of panel survey participation did not appear to have had an effect on sample characteristics (although younger participants tended to drop out disproportionately and part-time workers stayed in disproportionately). In particular, the similarity in transport characteristics between the samples suggests that panel survey participation did not bias the sample in this respect.

[Table 2 here]

The issue of attrition in the panel survey should be recognised. This applies to respondents who started to provide duration data at wave 2, but dropped out of the sample subsequently. None of the wave 2 respondents failed to respond at wave 3. Eight of the 247 panel participants did not respond to wave 4 and seven of these had not used Route 20 at wave 3 and hence are said to be right censored at wave 3 (week 15). Right censoring is a concern where it is not independent of the process being examined in which case it should be modelled jointly with the process (Jenkins 2004; 5-6). Given the small number of cases and no theoretical reason to believe that survey drop out was related to Route 20 use, no attempt is made to account for attrition in the following analysis.

Characteristics of the duration data

In this study, the duration data refers to the elapsed time spent by survey respondents without using Route 20. There is no left censoring of the duration data in this study, as measurement of usage of the Route 20 service was monitored as soon as it was introduced. It is assumed that individuals who have not used Route 20 may do so beyond the period of monitoring and therefore allowance is made for right censoring.

It is not only of interest to study the elapsed time until first use of Route 20, but also how long usage of Route 20 is sustained and if there is a subsequent period of non-usage. Duration models can be employed to study multiple spells, as shown by Hensher (1998) in considering the timing of automobile transactions. In our survey precise information was not available on how long Route 20 continued to be used after it was first used and it would be difficult to define in practically useful terms what is meant by continued use. As mentioned previously, dynamic choice models offer the possibility of analysing how persistence in Route 20 use changes over time, but this would be a separate analysis.

Duration analysis can be carried out on the basis of a continuous or discrete dependent variable. In this case the observations of duration times have been reported rounded to the nearest week and therefore the data relates to observations on a continuous random variable which are grouped (or interval censored). The most commonly available methods of duration modelling assume continuous observations but methods are also available for discrete observations. Jenkins (2004; 21) suggests that the smaller the ratio between the length of the intervals used for grouping and the typical duration length the more appropriate it is to use a continuous time specification. In this study we consider the implications of using both discrete and continuous specifications.

Descriptive statistics of duration data

A simple graphical plot of the number of new users of the Route 20 service over the survey period is shown in Figure 2. The number of new users is largest in the first week after introduction of the new service and tends to decline over time. ‘Spikes’ in new users occur in weeks 10 to 13 and 27. These weeks correspond to times when questionnaires were returned and may reflect some survey subjects indicating the week that they completed the survey as the first week they used Route 20, when actual first use occurred earlier.

[Figure 2 here]

Non-parametric duration analysis methods can be used to examine the duration data and gain an initial appreciation of its nature. In Table 3 statistics are presented on the Route 20 duration data based on grouping the duration times into three week periods. Grouping the data in this way aids quick inspection of how hazard rates vary through the measurement period. The hazard rate appears to be high initially and to generally decline over the period but with the previously mentioned spikes (which coincide with survey waves).

[Table 3 here]

External covariates

The external covariates that are tested in this study (for their effect on duration times) are shown in Table 4. The covariates are mainly personal and household characteristics of the residents, but also include three variables relating to bus access and travel times (which vary according to home location). These were calculated from the postcodes of the residents and the Transport Direct journey planner website (Transport Direct 2008). In the first part of Table 4, comparison is made of how the percentage of residents using Route 20 by the end of the survey period (March 2006) varies according to resident characteristics (these are initial characteristics at wave one of panel survey). This shows that use of the new bus service was higher for Broadfield residents, younger residents and residents without a driving licence or car in household.

[Table 4 here]

The second part of Table 4 provides statistics for covariates measured as continuous variables. It is noted that the personal characteristics covariates are all included as time non-varying covariates based on survey measurements at wave 1. There were a few cases where these characteristics changed through the survey period but these were so limited in number that it was not considered that it would provide significant additional explanation to attempt to include these as time-varying covariates. (Hensher (1997) considers a method for calculating continuous time-varying covariates from variables measured at discrete measurements.)

As well as personal characteristics, various attitudinal and behavioural measurements were obtained in the panel survey. These were not tested in the duration modelling since these would not strictly be exogenous covariates. For example, attitude towards buses as measured at wave 1 would be expected to influence hazard probability (i.e. positive attitude to buses increases hazard probability), but attitude to buses would be likely to be related to other unmeasured factors and any estimate of the effect of attitude to buses would capture both its true effect and unobserved heterogeneity. A solution to this is to instrument the attitude variable by regressing it against exogeneous variables and using the regression predicted values as variables in the duration model. This approach has not been been tested in this study, but it is a possible area for further work.

5. Modelling specifications

A number of modelling specification decisions needed to be made in the duration modelling. These are discussed next.

1. *Discrete or continuous time formulation*. As mentioned, given the grouped nature of the data it is more appropriate to use a discrete time formulation than a continuous time formulation[1]. Models were first developed for discrete time formulations, but comparison was also made with continuous time formulations and it was necessary to use continuous time formulations for the split population model described later in this section.

2. *External covariates*. Heterogeneity in duration times due to between-subjects differences is modelled through including external covariates in the hazard model. The effect of covariates on the hazard at time t is modelled through the proportional hazards parametric method which specifies the effect of external covariates to be multiplicative on the underlying hazard distribution (Bhat 2000). The other common method for modelling effect of external covariates is the accelerated failure time method. It is noted that the two methods are equivalent in the case of the Exponential and Weibull parametric distributions which are found to be appropriate distributions to use for the baseline hazard for the data in this study.

3. *Parameterisation of the baseline hazard*. The baseline hazard reflects the underlying duration dependence. Different baseline hazard distributions were tested and this is discussed in the results that follow. A non-parametric baseline hazard can be specified which has the advantage of avoiding forcing of an inappropriate distribution to the data. It is noted, however, that if it is wished to use the duration model to forecast durations beyond the data observation period then a parametric specification is required.

4. *Unobserved heterogeneity*. It may not be possible to control for all heterogeneity through covariates. With duration models the presence of unobserved variables tends to bias towards a negative duration dependence (Bhat 2000). Unobserved heterogeneity can be incorporated into a duration model through the inclusion of a random effect term (Bhat 2000). A parametric distribution or non-parametric distribution can be used for this term. This is

[1] The different formulation of hazard models for continuous time duration data and discrete time duration data are presented in Bhat (1996; 93-94), Steffensmeier and Jones (1997; 1424-7) and Jenkins (2004;13-24). In the continuous time formulation, it is hazard rate that is modelled, while in the discrete time formulation, it is hazard probability that is modelled.

tested in the results that follow.

5. *Spilt population*. In standard duration models it is assumed that all units will eventually experience the event of interest (in this case adopt Route 20 service). Another form of heterogeneity that can be envisaged is some units will never experience the event. In our case, there may be residents who are unable to use the Route 20 service (for example, due to physical limitations), or who are unwilling to use the service. A modified form of duration model exists which enables the assumption that all subjects will eventually experience the event to be relaxed. This is the split population duration (SPD) model, which has been introduced into econometrics by Schmidt and Witte (1984) and is tested in the results that follow. The derivation of the SPD model is set out in the Appendix to this paper.

The SPD model simultaneously considers the likelihood of the event occurring and the timing of the event. Two sets of coefficients are estimated: coefficients for the effect of covariates on the likelihood of the event occurring; and coefficients for the effect of covariates on the timing of the event occurring, conditional on the probability of the event occurring. The only other known applications of the SPD model in transport are by Hensher (1997) for modelling motorists' adoption of toll road and by Chang and Yeh (2007) for modelling motorcycle holding time.

6. Modelling results

Duration models were estimated using Stata (StataCorp 2005) and LIMDEP (Greene 2002)[2]. Results are presented and discussed in this section.

External covariates

An initial assessment was made of the bivariate associations between covariates and

---

[2] Discrete duration models without unobserved heterogeneity were estimated in Stata using the built-in routine cloglog. This routine estimates a discrete-time proportional hazard model based on a complementary log-log transformation of the interval hazard rate (Prentice and Gloeckler 1978). Discrete duration models with unobserved heterogeneity were estimated using estimation routines provided by Stephen Jenkins (University of Essex, UK) for Stata users: pgmhaz8 (Jenkins 1997) and hshaz (Jenkins 2004). Continuous duration models were estimated using Stata and Limdep built-in routines. SPD models for continuous data were estimated with Limdep.

duration time using a discrete-time model with a fully non-parametric baseline hazard specification. A fully non-parametric baseline hazard allows interval-specific baseline hazards to be estimated (which are independent of hazards in other intervals). The interval-specific baseline hazards can only be identified for those intervals during which the event is observed to occur. The data is therefore grouped into three week intervals for this purpose. Statistically significant associations (at 95% confidence level) were found for **Residential area** (living in Broadfield increases hazard), **Driving licence** (having driving licence decreases the hazard), **No car in household** (no car increases the hazard), **Two or more cars in household** (two cars decreases the hazard), **Bus pass** (having bus pass increases the hazard), **Walk access time to bus stop for Route 20** (greater access time decreases the hazard), **Reduction in walk access time to bus stop** and **Reduction in total time** (greater reduction in time increases the hazard).

A multivariate model was then estimated through a forward stepwise method. Statistically significant covariates (at 95% confidence level) were the two-level categorical variables **Residential area**, **No car in household**, **Bus pass** and **Commute to work**, and the continuous variable **Reduction in walk access time to bus stop**. Coefficient estimates and goodness-of-fit statistics are shown in Table 5. The implications of the covariate coefficients can be understood by noting that the effect on the hazard probability of a one unit change in a covariate, while controlling for other variables, is given by exp(β).

The coefficient estimates indicate that the signs of the covariate coefficients correspond to a priori expectations: the hazard rate increases if a resident lives in Broadfield (by factor of 3.4), lives in a household without a car (by factor of 6.7), has a bus pass (these provide free bus travel to those aged 60 and over or disabled who apply for the pass) (by factor of 3.1), commutes to work (by factor of 2.4) and experiences a decreased walking time to access bus services (by a factor of 1.06 for a reduction in walking access time of one minute). Specifically, the last item measures the reduction in walking time at the home end of the journey to access bus services to Gatwick Airport (which is taken as a representative destination location) resulting from the Route 20 service. The reason that residents living in

Three Bridges have a lower hazard rate than Broadfield residents can be suggested to be due to Three Bridges being better connected to destinations in the Crawley area by other public transport services and due to lower familiarity with the Fastway concept (Route 10 had previously been introduced to Broadfield).

[Table 5 here]

Parameterisation of the baseline hazard

An initial assessment of duration dependence compared the Weibull distribution with the Exponential distribution for a model with no covariates. This showed that the distribution parameter, p, was statistically significant with a value of 0.557 in the Weibull model (p is constrained to 1.0 in Exponential model). This indicates a monotonically decreasing hazard.

The fully non-parametric baseline hazard specification is compared to a Weibull baseline hazard specification in Table 5. With the five statistically significant covariates, the distribution parameter, p, is statistically significant with a value of 0.698 in the Weibull model. The results indicate that the covariate estimates are very similar for the two models. In the fully non-parametric model, the interval-specific hazard coefficients tend to decrease over time and it is concluded that the Weibull model represents a parsimonious specification of duration dependence (this is also supported by the adjusted likelihood ratio index being higher for Weibull model than fully non-parametric model).

Examination was made whether duration dependence varied within the sample of respondents based on the measured covariates. The Weibull parameter value, p, can be modelled as a function of covariates. The only covariate for which a statistically significant effect was found (at 95% confidence level) was the **Commute to work** two-level categorical covariate. It was only marginally statistically significant at 95% confidence level. For residents who commute to work, Weibull parameter value, p, is estimated to be 0.615 (decreasing hazard over time) and for residents who do not commute to work, p is estimated to be 1.09 (approximately constant hazard).

Examples of what these models imply are shown in Figures 3 and 4. Figure 3 plots two survival functions for the discrete Weibull model. Resident characteristics are assumed to be the same in the two cases, except with regard to **Reduction in walk access time to bus stop**. It is evident that in the case of a 15 minute reduction in access time the survival probability is considerably lower at any point in time. Figure 4 plots two survival functions for the discrete Weibull model where Weibull distribution parameter, p, was modelled as a function of **Commute to work**. Resident characteristics are assumed to be the same in the two cases, except with regard to whether the resident commutes or not. It is evident that the commuter has a lower probability of not using Route 20 at any point in time after its introduction. However, the rate over time at which the survival probability decreases appears to be constant for the non-commuter, whilst it decreases substantially at first for the commuter but more slowly subsequently. This is a consequence of the differing values for the Weibull, p, parameter for commuters and non-commuters.

[Figure 3 here]

[Figure 4 here]

Unobserved heterogeneity

It is important to test for the presence of unobserved heterogeneity in duration models and Bhat (2000) emphasises the need for considering parametric and non-parametric unobserved heterogeneity specifications. The Gamma (parametric) mixing distribution and discrete (non-parametric) mixing distribution with finite number of support points have been tested with both the discrete Weibull and discrete fully non-parametric baseline hazard models. The estimated parameters for unobserved heterogeneity were found to be statistically significant in duration models without covariates, but not found to be statistically significant after including the five covariates. One result is shown in Table 5. There was no evidence that ignoring heterogeneity results in attenuation of parameters (covariates or

Weibull p parameter) which has been found to be an issue with some data sets.

Continuous time formulation

A continuous specification was compared to a discrete specification. With a Weibull parametric distribution (and the same covariates) the discrete and continuous models are compared in Table 5. Covariate parameter estimates are found to be very similar. A difference, however, is that the Weibull, p, parameter is estimated to be 0.698 in discrete model and 0.897 in continuous model.

Split population model

A final form of heterogeneity that we wished to test is whether there are some residents who will never use Route 20. The split population duration (SPD) model can be used to test this and was modelled using Limdep[3].

In Table 5 estimation results are shown for a SPD model (continuous Weibull parametric specification for duration part of model and logit model for event occurrence part of model). The covariates included in the event occurrence part of the model were selected based on the separate estimation of a logit model for probability of using Route 20 by the end of survey period. Statistically significant variables at the 95% confidence level were **Residential area**, **No car in household**, **Bus pass** and **Reduction in walk access time to bus stop**.

Comparison of the standard duration model and the SPD model shows that model fits are very similar with the two models having the same adjusted likelihood ratio index of 0.165. The split parameter is estimated to be 0.638 which implies that about 36% of residents will never use Route 20. The statistical significance of the split parameter is marginal (it is not found to be different from one at the 95% confidence level). Nevertheless, the SPD model is informative in indicating the differing role that covariates may have in influencing whether

[3] The only known estimation routine for fitting a SPD model with a discrete time formulation is *spsurv* which has been written for Stata, but this does not allow the probability of event to be modelled with covariates (Lambert 2007). In Limdep Version 8.0 the facility is provided to fit a SPD model with a continuous time formulation and with both the event occurrence and timing parts of the model able to be modelled with covariates (Greene 2002; E27:25-27).

Route 20 is used and when it is used. The Weibull, p, parameter is estimated to be 1.067 for SPD model and 0.897 in standard duration model. This implies that, after accounting for both the effect of covariates on hazard and the possibility that some subjects will never use the new service, there is no longer a negative duration dependence.

In the SPD model it is found that the probability of ever using Route 20 is higher for residents who live in a household without a car and who experience a decreased walking time to access bus services. The other two covariates are not statistically significant. The parameter estimates for the hazard model indicate that the hazard rate is higher if a resident lives in Broadfield, lives in a household without a car, has a bus pass and commutes to work. A decreased walking time to access bus services is not statistically significant for hazard rate after accounting for probability of using Route 20.

After accounting for the probability of the event occurrence, the hazard model parameter estimates in the SPD model are quite different to their values in the standard duration model. The size of the effect on hazard rate of living in a household without a car is lower in the SPD model having taken in account the probability of using Route 20. In contrast, the size of the effect on hazard rate of living in Broadfield is higher in the SPD model having taken in account the probability of using Route 20.

The predictive implications of the standard duration model and the SPD model are now examined. First, within-sample prediction is carried out and a comparison is made of the predicted survival times of the two models with the observed survival times for the analysis sample. Survival functions estimated for the models calculate the probability of survival to time t (i.e. time spent without using Route 20). From the survival function the *survival median time* t is calculated as the time for which survival probability, S(t), equals 0.5. This represents the time t when there is a 50% probability of survival and 50% probability of not survival and therefore can be taken as the *predicted* survival time.

To calculate survival median times for the SPD model, it needs to be recognised that some of the sample are predicted to never use Route 20. The procedure used is the same as that by Cushing Daniels (2005). First, the probability of the event occurring is calculated for all sample respondents (from the event occurrence (logit) part of the SPD model) and respondents are ranked in decreasing probability of experiencing the event. Then, noting that the SPD model predicts that 63.8% of the sample will eventually use Route 20, only the respondents with the 63.8% highest probabilities are selected and median survival times are calculated for these respondents (from the timing (hazard) part of the SPD model). The other respondents are assumed to not ever experience the event.

Figure 5 compares the cumulative number of residents observed to have used Route 20 (from the survey data) with the number expected to have used Route 20 based on the predicted survival times from the standard duration model and the SPD model. It shows that both duration models consistently under-predict the number of users of Route 20 within the 27 weeks of the survey period. This is particularly marked for the standard duration model. However, looking slightly beyond the 27 week period, the prediction lines cross at 30 weeks after which the standard duration model predicts a higher number of Route 20 users than the SPD model and would probably make an over-prediction of the number of Route 20 users.

Many of the observations in the analysis sample were right censored (70%) and it is important to look at the predictions from the standard duration model and the SPD model for the full sample of respondents. This is shown in Figure 6. The fundamental differences in the two duration models can be seen. Only 63.8% of sample respondents are predicted to use Route 20 with the SPD model. All respondents who will use Route 20 are predicted to do so after about 200 weeks with the SPD model, whilst it is 600 weeks (over 10 years!) before all respondents are predicted to use Route 20 with the standard duration model.

When considering the predictive capabilities of duration models it is important to take into account their performance across a broad time frame. In comparison with the observed data, the SPD model performs better than the standard duration model in the short run, but less well in the medium run. It is uncertain which model performs better in the long run. A longer survey period would have helped to provide evidence for this.

[Figure 5 here]

[Figure 6 here]

The out-of-sample predictive implications of the two duration models can be considered using the survival functions. Figures 7 and 8 illustrate this by showing the survival functions for two different hypothetical residents. In Figure 7 the survival functions are similar for the two models. This resident has a very high probability of using Route 20 according to the SPD model (probability = 0.95) and with both models is predicted to not take long to start using Route 20 (median survival time of less than four weeks). In Figure 8 the survival functions are very different for the two models. This resident has a low probability of using Route 20 according to the SPD model (probability = 0.15). With the standard duration model the resident is more likely to not have used Route 20 after 50 weeks than to have used it. With the SPD model there are different implications. In the unlikely event of using Route 20, it is predicted to be used much sooner (median survival time of nine weeks).

[Figure 7 here]

[Figure 8 here]

7. Summary and conclusions

Before considering the substantive results, it is important to reflect on what has been learnt about model specification. A discrete-time formulation is appropriate for the data obtained in this study (interval data), but it was found that covariate coefficient estimates were very similar when a continuous-time formulation was used. There was a small difference in the duration dependence parameter between the two model formulations. After controlling for covariates, it was found with this data set there was no unobserved heterogeneity (whether unobserved heterogeneity was specified as parametric or non-parametric).

The first research hypothesis that the hazard probability of using Route 20 declines over the measurement period is found to be supported by the data, although after controlling for covariates, the effect is slight. The second hypothesis that the hazard probability of using Route 20 increases for those residents gaining greater accessibility benefits from Route 20 is found to be supported by the data, although the size of effect is modest compared to that of other between-resident differences. Increased hazard probability is found (in descending order of magnitude) for residents not having a household car, living in Broadfleld, having a bus pass and being a commuter.

More sophisticated attempts at recognising heterogeneity were tested. There was an indication that duration dependence differed within the sample of respondents based on whether respondents were commuters or not. This was marginally statistically significant at 95% confidence level. A negative duration dependence is found for those residents that commute to work and a constant duration dependence for those that do not commute to work. An explanation is that where Route 20 offered a useful alternative to get to work it is likely to have been tried soon after the service was introduced. It becomes increasingly unlikely to be used after this, as it does not suit the commute journeys of remaining residents.

Allowance for the possibility that there existed a proportion of the sample that would never use Route 20 was tested using a SPD model. In a SPD model a split parameter lower than one indicates the existence of subjects who will never experience the event of interest. The split parameter was estimated to be 0.638, but a difference from one could not be supported at the 95% confidence level. Nevertheless, the SPD model is informative in indicating that some covariates are more important in influencing whether Route 20 is used than when it is used. These covariates are living in a household without a car and a decreased walking time to access bus services. A larger data set and longer survey period would have helped to provide stronger evidence for whether the SPD model is an appropriate model in this context. In principle, the SPD model is an attractive model that offers combined information on how many people will use a public transport service and when they will start to use it.

It is recommended that the opportunity is taken in future when new or modified public transport services are introduced to obtain duration data for the elapsed time taken until people start to use the new service. Different methods of collecting duration data to that in this study should be considered. If fare payments are made using Smartcards then these could offer possibilities for collecting duration data. Smartcards enable the monitoring of usage of a service over time and would allow not only first usage of a new service to be identified but also subsequent pattern of usage. Another approach is to distribute introductory discount (or free) offers to use a new service to relevant members of the population (residents, employees) and require them to submit a form with their address and other details when they first use the service. The requirement in both cases would be to not only obtain data for those people that eventually use the service, but also those people that do not use the service. This could be achieved by selecting a random sample of the population for the study, requesting them to provide personal, household and travel details and providing them with the discount offer or Smartcard with which their usage of the new service can be monitored.

Marketing and service implications can be drawn from the duration modelling results. For those groups within the population more likely to have responded quickly in using Route 20 (e.g. Broadfield residents, commuters), marketing can be aimed at reinforcing the value of the new service to these groups. For those groups less likely to have responded quickly in using Route 20, service modifications can be considered to provide a more attractive transport option and targeted marketing can be considered to publicise and promote the service. For example, the slower adoption of the new service by residents in Three Bridges suggests that advance marketing of the new bus service to those living close to the service in this area could have been worthwhile.

A major concern raised at the start of the paper is the current inability to forecast patronage growth. Predicting the overall number of users of a service, at any time after it is introduced or modified, based on predictions of the numbers of people who will have started to use the service will result in over-predictions. Some users will not persist in using a public transport service after initially trying it. This is shown in Chatterjee and Ma (2007). Modelling residents' dynamic travel choices (based on observations at panel occasions) does not fully utilise the duration times obtained, but allows persistence in usage to be analysed. This is a more appropriate approach for developing a patronage forecasting tool and is the subject of separate modelling analysis as reported in Chatterjee (2008).

The contribution of this study has been to apply duration modelling to timing of adoption of a new public transport option. Analysing durations enables the full information available on duration times from the panel survey to be used. Alternative modelling specifications have been used to examine heterogeneity in responses amongst the population of residents in the study area. This has allowed a better understanding to be gained on how people respond to a new public transport service and it is considered that many of the insights are likely to be applicable to other contexts. The analysis in this paper has examined the effect of objective factors on durations and it would be valuable to extend the analysis to examine the effect of subjective factors (e.g. awareness, attitudes, habit).

8. Acknowledgements

The research reported in this paper has been funded by a grant from the UK Engineering and Physical Sciences Research Council (EPSRC).

9. References

Bhat CR (1996) A hazard-based duration model of shopping activity with nonparametric baseline specification and nonparametric control for unobserved heterogeneity. Transp Res Part B 30:189-207

Bhat CR (2000) Duration models. In Hensher DA, Button KJ (eds) Handbook of transport modelling. Elsevier, Oxford 91-111

Box-Steffensmeier JM, Jones BS (1997) Time is of the essence: event history models in political science. Am J Political Sci 41:1414-1461

Chang H-L, Yeh T-H (2007) Exploratory analysis of motorcycle holding time heterogeneity using a split-population duration model. Transp Res Part A 41:587-596

Chatterjee K, Ma K (2007) Modelling the timing of user responses to a new urban public transport service: application of duration modelling. Transp Res Record 2010:62-72

Chatterjee (2008) Towards the dynamic modelling of travel demand. Unpublished paper presented at the 40th Annual Conference of the Universities' Transport Study Group, Portsmouth

Cox D, Oakes D (1984) Analysis of survival data. Chapman and Hall/CRC, Boca Raton, FL

Cushing-Daniels B (2005) Even the errors discriminate: how the split-population model of criminal recidivism makes justice even less colorblind. Rev Black Political Econ 33:25-40

DfT (2003) Major scheme appraisal in local transport plans part 1: detailed guidance on public transport and highways schemes. Department for Transport, London.

Douglas N (2003) Patronage ramp-up factors for new rail services. Unpublished paper. Available via Chartered Institute in Logistics and Transport in New Zealand. http://www.cilt.co.nz. Cited 30 May 2006

Fastway (2008) Fastway homepage. Available via West Sussex County Council. http://www.westsussex.gov.uk/ccm/navigation/roads-and-transport/public-transport/fastway/. Cited 9 May 2008

Greene WH (2002) LIMDEP version 8.0: econometric modelling guide volume 2. Econometric Software, Plainview, NY

GSR (2004) The Magenta book: guidance notes for policy evaluation and analysis. Background paper 5: what is sampling? Government Social Research Unit, London. Available via http://www.gsr.gov.uk/professional_guidance/magenta_book/index.asp. Cited 9 May 2008

Hensher DA (1997) The timing of change: discrete and continuous time panels in transportation. In Golob TF, Kitamura R, Long L (eds) Panels for transportation planning. Kluwer, Boston 305-319

Hensher DA (1998) The timing of change for automobile transactions: a competing risk multispell specification. In Ortuzar JD, Hensher DA, Jara-Diaz S (eds) Travel behaviour research: updating the state of play. Pergamon, Oxford 487-506

Hensher DA, Mannering FL (1994) Hazard-based duration models and their application to transport analysis. Transp Rev 14:63-82

Lambert PC (2007) Modeling of the cure fraction in survival studies. The Stata Journal 7:1-25

Le CT (1997) Applied survival analysis. Wiley, New York

Jenkins SP (1997) Discrete time proportional hazards regresion. Stata Technical Bulletin 39:22-32

Jenkins SP (2004) Survival analysis. Unpublished manuscript. Institute for Social and Economic Research, University of Essex, UK. Available via Institute for Social and Economic Research. http://www.iser.essex.ac.uk/teaching/degree/stephenj/ec968/pdfs/ec968lnotesv6.pdf. Cited 9 May 2008

Petersen T (1991) The statistical analysis of event histories. Social Methods Res 19:270-323

Prentice RL, Gloeckler L (1978) Regression analysis of grouped survival data with application to breast cancer data. Biometrics 34:57-67

Schmidt P, Witte AD (1989) Predicting criminal recidivism using ‘split population’ survival time models. J Econ 40:141-159

StataCorp (2005) Stata Statisitcial Software: Release 9. StataCorp LP, College Station, TX.

Transport Direct (2008) Transport Direct homepage. Available via Transport Direct. http://www.transportdirect.info. Cited 9 May 2008

Washington SP, Karlaftis MG, Mannering FL (2003) Statistical and econometric methods for transportation data analysis. Chapman and Hall/CRC, Boca Raton 217-237

Yamaguchi K (1991) Event history analysis. Sage, Newbury Park

10. Appendix – the Split Population Duration (SPD) model

The split population duration (SPD) model enables the assumption to be relaxed that all subjects will eventually experience the event being modelled. It has been introduced into econometrics by Schmidt and Witte (1984).

The specification of the SPD model is introduced by first noting that the general likelihood function for the standard duration model specification is the following:

$$L = \prod_{i=1}^{N} \left[f(t;x) \right]^{C_i} \left[S(t;x) \right]^{1-C_i} \tag{1}$$

where N is the number of observations, i is index of observations, C_i is an indicator variable that equals 1 when adoption of Route 20 occurs by time t and 0 otherwise, f(t;x) is the probability density function and S(t;x) is survival function.

The probability density function, f(t;x), contributes to the likelihood for observations for which event occurs by time t and the survival function, S(t;x), contributes to the likelihood for observations for which event does not occur by censoring point time t.

With the SPD model a latent variable, U_i, is defined which equals 1 for observations that will eventually experience event (use Route 20 in this case) and equals 0 for those observations which will never experience event. The probability, $U_i = 1$, is defined as δ which represents the split population parameter and if this is less than 1 then some of the censored observations will never experience the event.

For observations that experience the event ($C_i = 1$) it is implied that $U_i = 1$ and the density function is:

$$\Pr(U_i = 1)\Pr(T_i \le t | U_i = 1) = \delta f(t; x, U = 1) \tag{2}$$

For censored observations ($C_i = 0$) there are two possibilities: that the event will never occur or it is right censored. The density function is:

$$\Pr(U_i = 0) + \Pr(U_i = 1)\Pr(T_i > t | U_i = 1) = (1-\delta) + \delta S(t; x, U = 1) \tag{3}$$

The likelihood function for SPD model is therefore a combination of the two functions (2) and (3) and is fully specified based on the parametric specification used (Weibull in this case).

$$L = \prod_{i=1}^{N} \left[\delta f(t; x, U = 1)\right]^{C_i} \left[1 - \delta + \delta S(t; x, U = 1)\right]^{1-C_i} \tag{4}$$

The probability of δ (the incidence portion of the model) is typically modelled as a logit or probit and can include exogeneous variables which may be the same as or different from those included in the duration model (the timing portion of the model). The model is estimated using the censoring indicator (observation of whether event occurred or not) as dependent variable.

$$\delta = F(\alpha Z) = 1/\left[1 + \exp(\alpha Z)\right] \tag{5}$$

In the SPD model the coefficients for the effect of exogeneous variables on the incidence of the event occurring are estimated as well as coefficients for the effect of exogeneous variables on the timing of the event, conditional on the probability that the event occurs. Different exogeneous variables can be included in the incidence and timing parts of the model. The SPD model estimates the split parameter, δ, which is the estimated mean probability of cases experiencing the event of interest. This parameter allows the test to be made of whether relaxing the assumption that every observation will experience the event is necessary. If $\delta = 1$ for all observations the likelihood reduces to a standard duration model with right censoring.

Figures

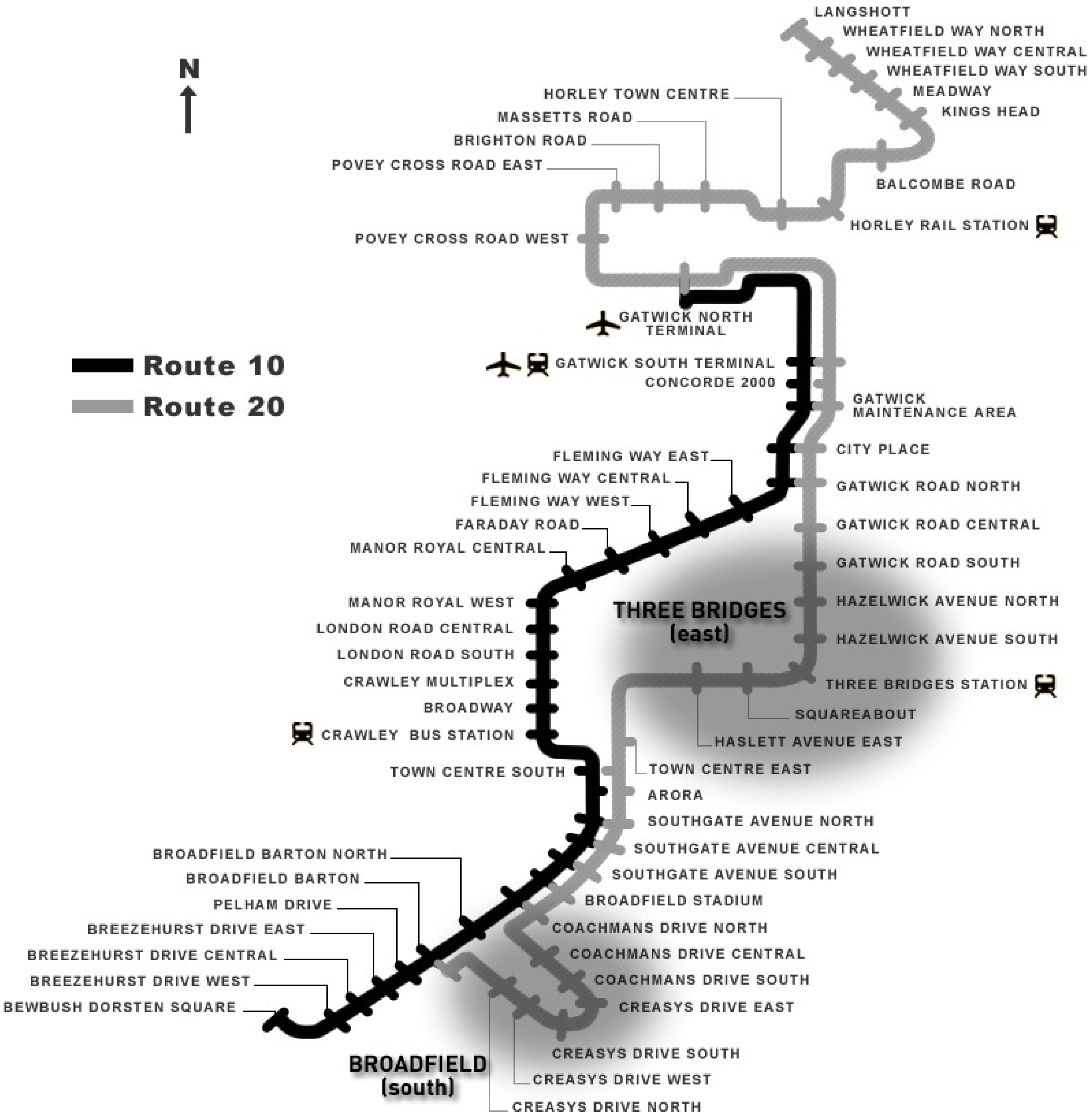


Figure 1. Fastway Routes (source: Fastway 2008)

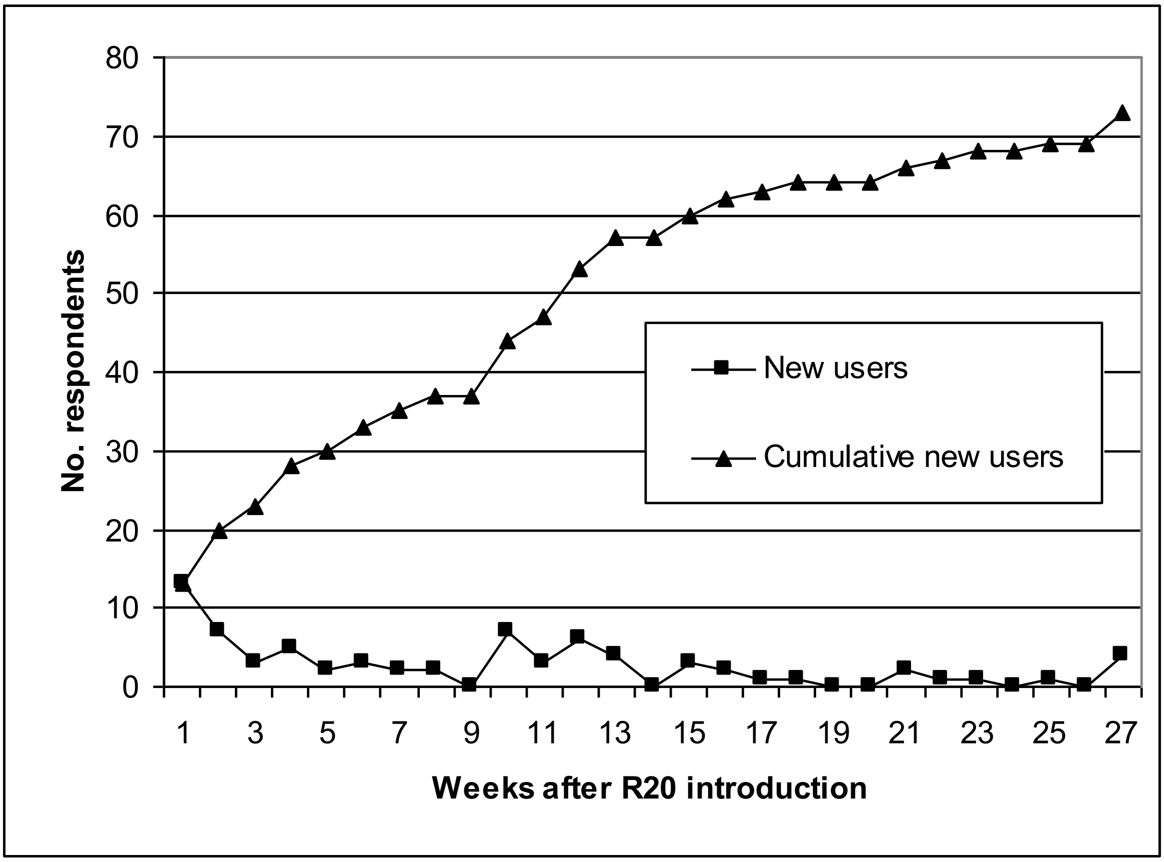
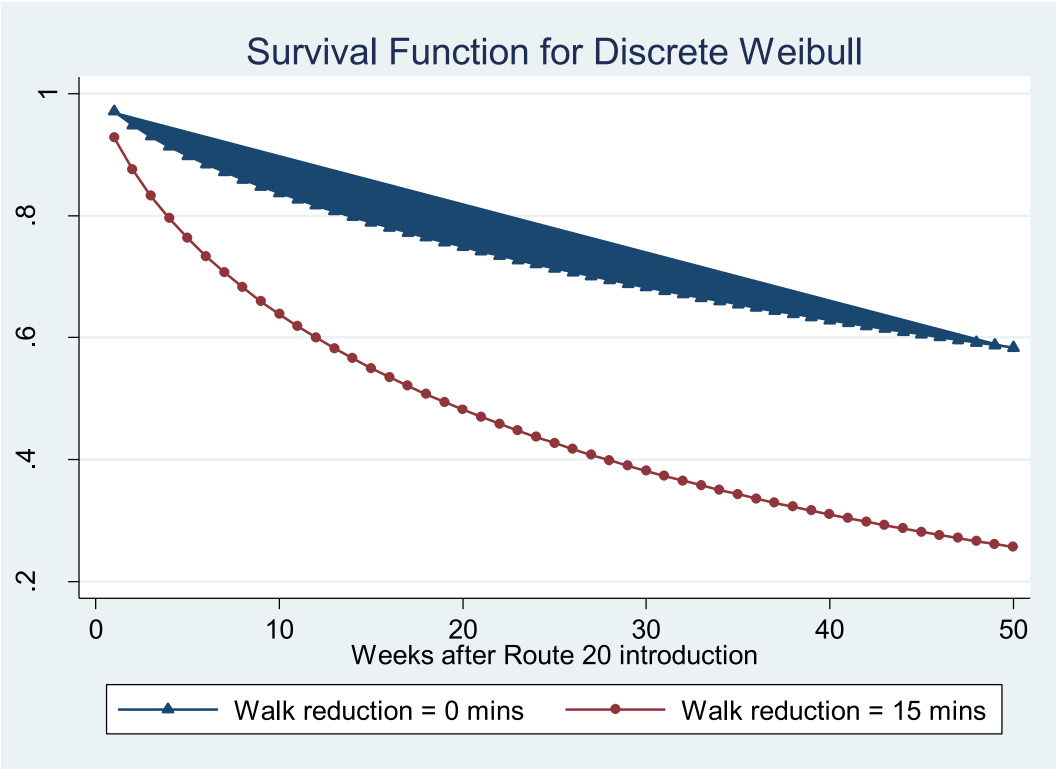
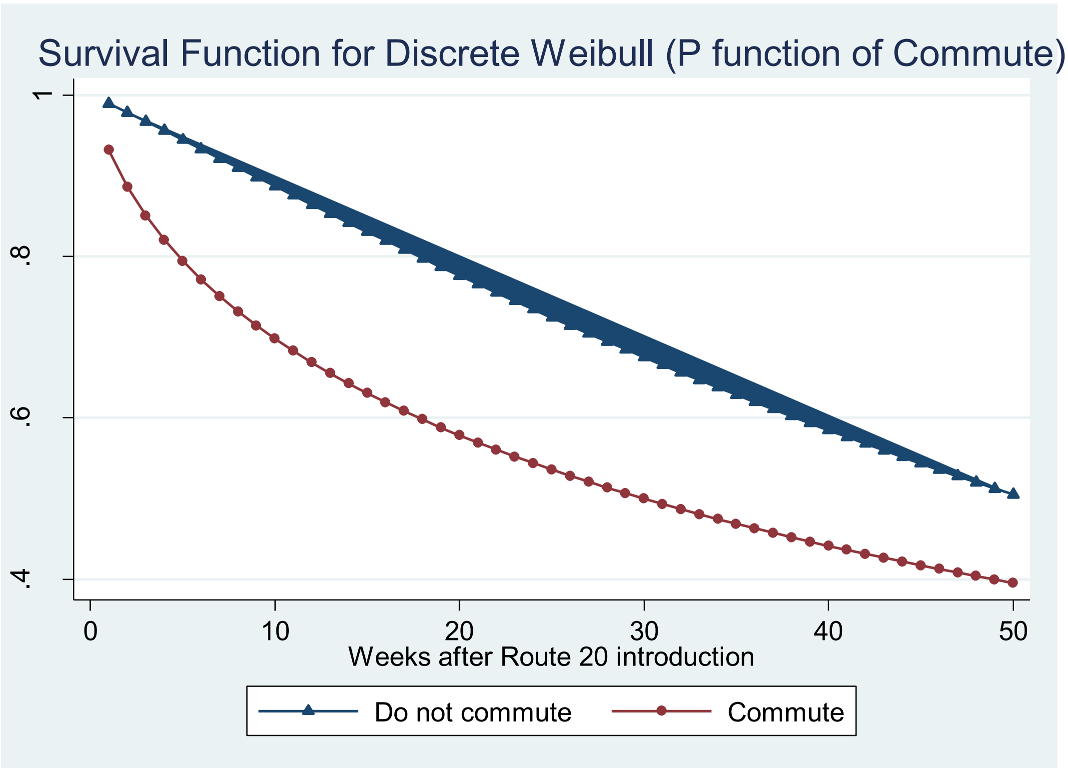


Figure 2. Number of New Users of Route 20 in the Weeks after its Introduction



† Resident characteristics – live in Broadfield, commutes, have car and do not have bus pass

Figure 3. Survival Function Comparisons for Reduction in Walk Access Time



† Resident characteristics – live in Broadfield, have car, do not have bus pass and walk reduction of 10 mins

Figure 4. Survival Function Comparisons for Commute or Not Commute

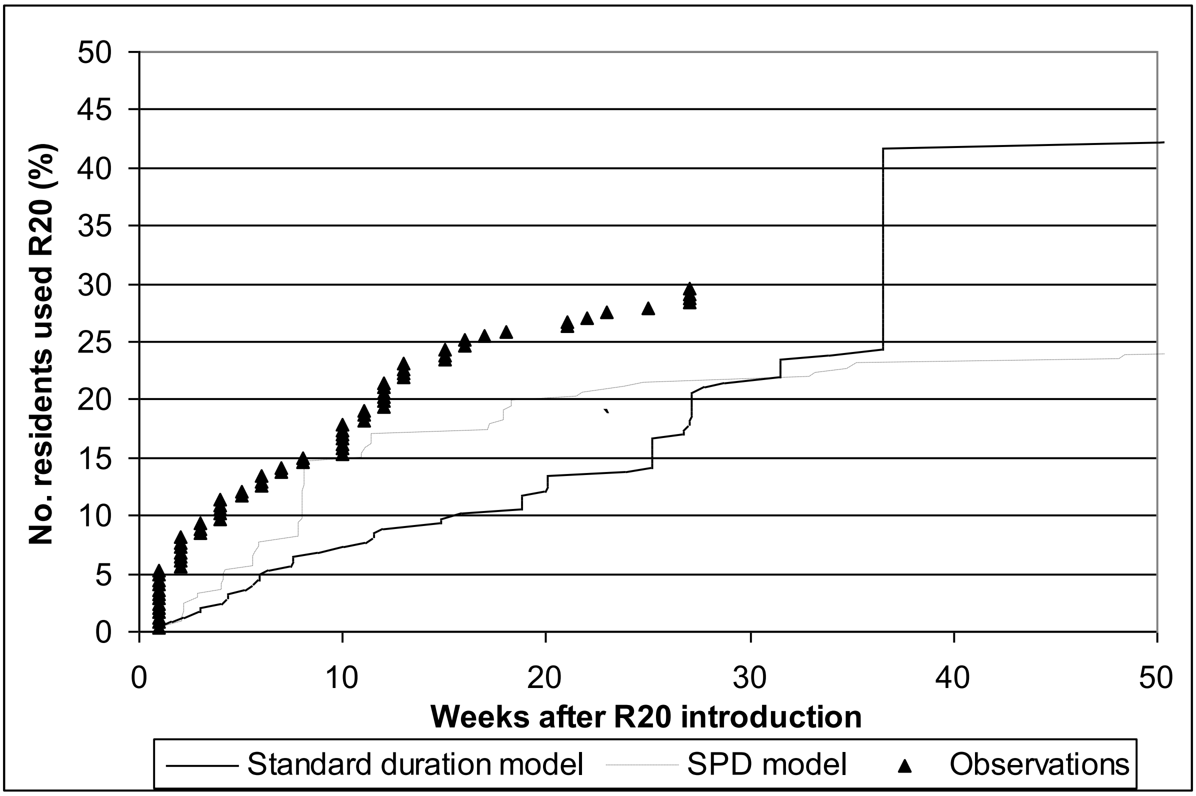


Figure 5. Comparison of Observed Number of Users of Route 20 with Predictions from Standard Duration Model and SPD Model

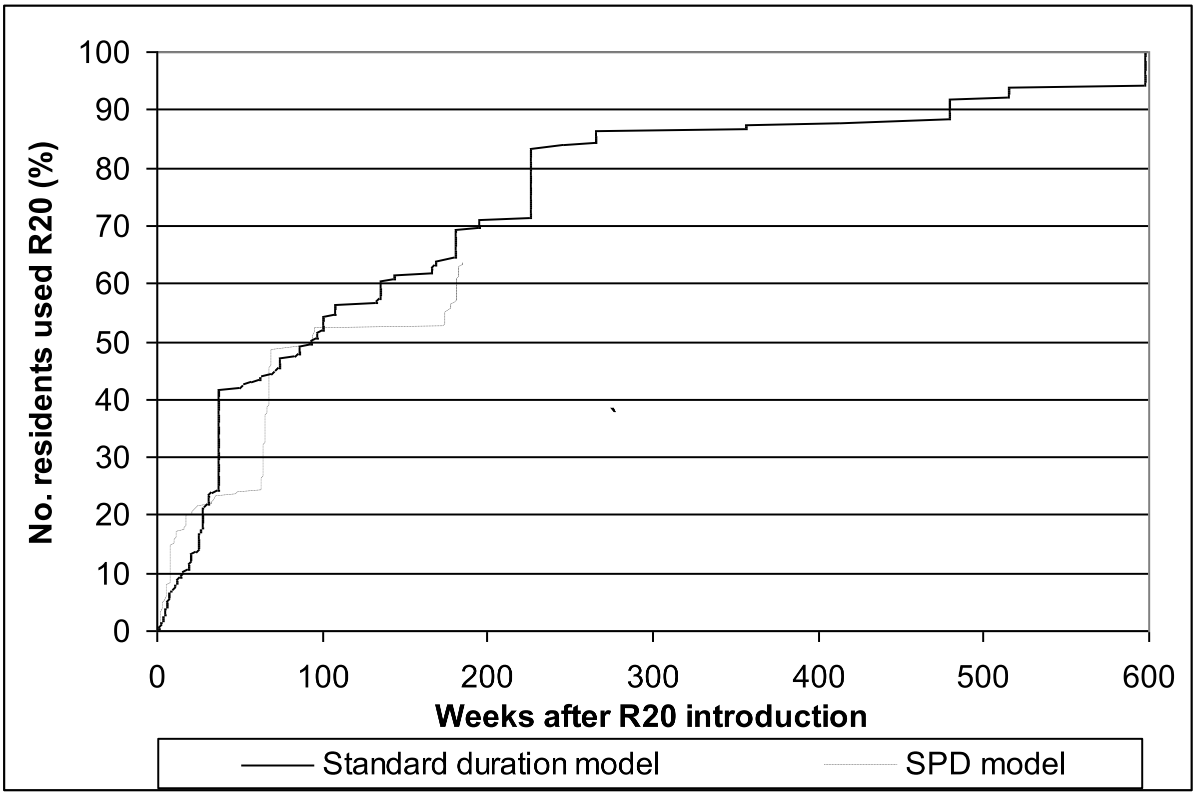
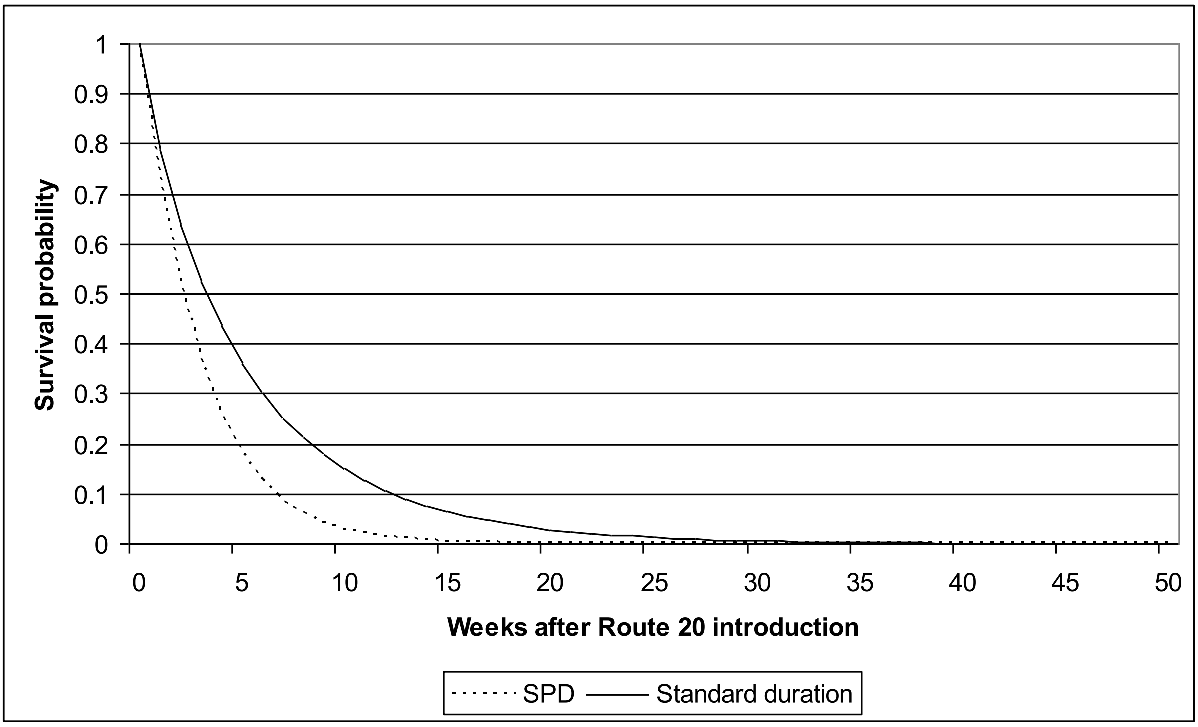
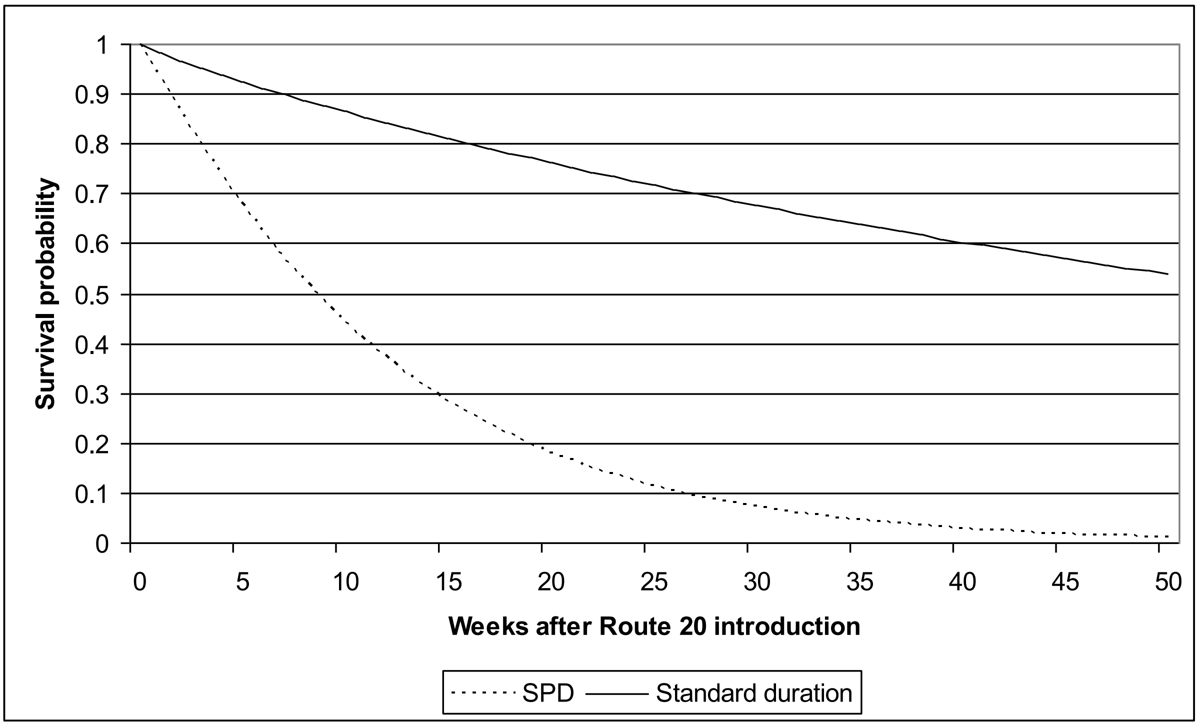


Figure 6. Comparison of Predicted Number of Users of Route 20 from Standard Duration Model and SPD Model for Full Sample



† Resident characteristics – Broadfield, no car, no bus pass, commute, walk reduction of 10 mins (SPD model: probability of using Route 20 = 0.95)

Figure 7. Comparison of Predicted Survival Times for Standard Duration Model and SPD Model



† Resident characteristics – Broadfield, car, no bus pass, commute, walk reduction of 0 mins (SPD model: probability of using Route 20 = 0.15)

Figure 8. Comparison of Predicted Survival Times for Standard Duration Model and SPD Model

Tables

Table 1. Characteristics of Broadfield South and Three Bridges electoral wards

Characteristic	Broadfield South	Three Bridges	Crawley
Location	Edge of town neighbourhood	Inner town neighbourhood	-
Public transport (prior to Fastway Routes 10 and 20 introduction)	One bus service to town centre	National rail station and various bus services (mostly on boundary of neighbourhood)	-
Index of Multiple Deprivation[1]	3 of the 4 ward sub-areas are ranked in the top decile of deprived sub-areas in the county of West Sussex	None of the sub-areas within ward are ranked in the top decile of deprived sub-areas in the county of West Sussex	7 sub-areas within town are ranked in the top decile of deprived sub-areas in the county of West Sussex
Percentage of population aged 65 and over[2]	5.2	20.7	14.7
Mode share percentages for travel to work[2]			
Car	69.6	60.5	67.5
Train	3.1	8.0	6.2
Bus	11.7	2.1	6.3
Walking	4.8	14.4	7.8
Percentage with distance to work less than 2km[2]	7.1	38.8	19.3
Percentage of households without car[2]	22.4	22.1	20.4

[1] Office of the Deputy Prime Minister (ODPM) English Index of Multiple Deprivation 2004 which is a measure of multiple deprivation at the small area level and is an index based on seven domains of deprivation.
[2] From 2001 Census.

Table 2. Survey Sample Characteristics

Characteristic	Crawley population (from Census 2001) (%)	Wave 1 Respondents (N=554) (%)	Analysis Sample Respondents (N=247) (%)
Female	51	55	55
Aged under 35 (and >16)	34	27	21
Aged 65 and over	19	19	17
Full-time employed	Not known	52	52
Part-time employed	Not known	13	18
Households without car	20	9	10
Used Route 10 service	Not applicable	29	30
Intending to use new Fastway service	Not applicable	27	27

Table 3. Discrete Period Durations and Sample Hazard

Period	Time interval (weeks)	'Failures'	Right censored	No. 'at risk' at beginning of period	Discrete period hazard †	Standard error of hazard †
1	0 - 3	23	0	247	0.0326	0.0068
2	4 - 6	10	0	224	0.0152	0.0048
3	7 - 9	4	0	214	0.0063	0.0031
4	10 - 12	16	0	210	0.0264	0.0066
5	13 - 15	7	7	194	0.0125	0.0047
6	16 - 18	4	0	180	0.0075	0.0037
7	19 - 21	2	0	176	0.0038	0.0027
8	22 - 24	2	0	174	0.0039	0.0027
9	25 - 27	5	167	172	0.0194	0.0087

† Estimated by lifetable method with actuarial adjustment for number of subjects at risk

Table 4. Resident Characteristics and Variation in Route 20

Resident Characteristic	Categories	Total number of respondents	Percentage used Route 20
Total analysis sample		247	30
Residential area	0 = Three Bridges	149	16
	1 = Broadfield	98	50
Gender	0 = Female	136	32
	1 = Male	111	27
Age (categorical)	0 = Under 25	20	45
	1 = 25-34	33	30
	2 = 35-44	56	32
	3 = 45-54	51	29
	4 = 55-64	46	20
	5 = 65 and over	41	29
New resident	0 = Not new resident	234	30
	1 = Less than one year	13	23
Live with spouse	0 = No	113	36
	1 = Yes	134	24
Driving licence	0 = No	34	53
	1 = Yes	213	26
Full-time employed	0 = No	119	31
	1 = Yes	128	28
Part-time employed	0 = No	202	28
	1 = Yes	45	36
Retired	0 = No	201	31
	1 = Yes	46	22
Children in household	0 = No	199	29
	1 = Yes	48	33
Cars in household	0 = 0 car in household	25	84
	1 = 1 car in household	120	25
	2 = 2 or more cars in household	102	22
No car in household	0 = No	222	23
	1 = Yes	25	84
Two or more cars in household	0 = No	145	35
	1 = Yes	102	22
Bus pass	0 = No	218	27
	1 = Yes	29	52
Commute to work	0 = No	74	27
	1 = Yes	173	31
Job change in last three months	0 = No	237	29
	1 = Yes	10	50

Resident Characteristic	Min-imum	Max-imum	Mean (standard deviation)
Age	17	82	47.9 (15.60)
Walk access time to bus stop for Route 20 bus to Gatwick (minutes)	1	15	6.91 (2.94)
Reduction in walk access time to bus stop for bus to Gatwick after Route 20 introduced (minutes)	0	16	6.01 (4.62)
Reduction in total time to travel to Gatwick by bus after Route 20 introduced (minutes)	0	28	11.5 (5.96)

University of the West of England BRISTOL

This is a pre-publication version of the following article:
Chatterjee, K. and Ma, K. (2009). Time taken for residents to adopt a new public transport service: Examining heterogeneity through duration modelling. *Transportation*, 36(1),1-25.

Centre for Transport & Society

Table 5. Duration Model Results

Model	**Discrete non-parametric**	**Discrete Weibull**	**Discrete Weibull p is fn(Commute to work)**	**Discrete Weibull with unobs. het.**	**Continuous Weibull**	**Split population /continuous Weibull**
Parameters and model performance	Estimated parameter β (z statistic)	Estimated parameter β (z statistic)	Estimated parameter β (z statistic)	Estimated parameter β (z statistic)	Estimated parameter β (z statistic)	Estimated parameter β (z statistic)
Hazard model covariates	Five	Five	Five	Five	Five	Five
Area (0=Three Bridges, 1=Broadfield)	1.221 (4.55)**	1.204 (4.48)**	1.232 (4.60)**	1.204 (1.41)	1.238 (4.63)**	2.226 (4.92)**
No household car (0=1 or more cars, 1=no household car)	1.897 (6.14)**	1.862 (6.03)**	1.869 (6.17)**	1.862 (5.14)**	1.906 (6.16)**	1.414 (3.07)**
Bus Pass (0=none, 1=bus pass)	1.143 (3.41)**	1.111 (3.33)**	1.060 (3.23)**	1.111 (3.03)**	1.152 (3.44)**	0.712 (1.69)
Commute to work (0=no, 1=yes)	0.887 (2.66)**	0.870 (2.64)**	1.907 (2.97)**	0.870 (4.68)**	0.873 (2.64)**	1.056 (2.66)**
Reduction in walk access time to bus stop for bus to Gatwick after Route 20 introduced (minutes)	0.0627 (2.03)*	0.0617 (2.00)*	0.0688 (2.21)*	0.0617 (1.30)	0.0663 (2.14)*	0..00577 (0.122)
Baseline (duration distribution) parameters	Nine	Two	Three	Two	Two	Two
Constant	-	-5.594 (-10.73)**	-6.481 (-8.85)**	-5.594 (-15.18)**	-6.103 (-10.82)**	-5.939 (-7.06)**
p (distribution parameter)	-	0.698$^{\wedge\wedge}$	1.094 (not commute) 0.615$^{\wedge}$ (commute)	0.698$^{\wedge\wedge}$	0.897	1.067
Unobserved heterogeneity parameters				One		
Gamma				3.67e-08 (0.03)		
Logit model covariates						Five
Constant						0.0768 (0.05)
δ (split parameter)						0.638
Model performance statistics						

Log Likelihood with covariates (LLC)	-323.7	-330.1	-327.9	-330.1	-207.8	-202.8
Log Likelihood with baseline parameters (LLB)	-364.7	-372.7	-372.7	-372.7	-252.4	-252.4
Log Likelihood at zero (LL0)	-390.6	-390.6	-390.6	-390.6	-257.3	-257.3
Adjusted likelihood ratio index = 1 – { (LLC – no. parameters in model) / LL0 }	0.135	0.137	0.140	0.134	0.165	0.165

** indicates that β significantly different than 0 at α = 0.01, * indicates that β significantly different than 0 at α = 0.05
^^ indicates that p significantly different than 1 at α = 0.01, ^ indicates that p significantly different than 1 at α = 0.05
# LLB is log likelihood with constant and adjusted likelihood ratio index is calculated with respect to LLB not LL0