HAZARD BASED MODELLING OF VEHICLE SELECTION IN CARSHARING SYSTEMS

Sisi Jian*, Kasun Wijayaratna, Taha Rashidi, Vinayak Dixit
School of Civil and Environmental Engineering, University of New South Wales
Sydney NSW 2052, Australia

*Corresponding author
Email: s.jian@unsw.edu.au

A Paper Submitted for Presentation at the 94th Annual Meeting (2015) of the Transportation Research Board

Total words = 5802 words text + 6figures/tables x 250 words (each) = 7302 words

Submission Date: 1st August, 2014

ABSTRACT:

Carsharing, as an alternative to private vehicle ownership, has spread worldwide in recent years due to its potential of reducing congestion, improving auto utilization rate and limiting the environmental impact of emissions release. To determine the most efficient allocation of resources within a carsharing programme, it is critical to understand what factors affect the users' behaviour when selecting vehicles which have been little attended to in the literature. This study attempts to investigate the importance of users' attributes and fleet characteristics on choice set formation behaviour in selecting vehicles using a Spatial Hazard Based Model (SHBM). In the SHBM model, "distance to the carsharing facility (vehicle pod)" is considered as the prospective decision criteria that carsharing users follow when selecting a vehicle among a set of alternatives This variable is analogous to the duration in a conventional hazard-based model Other factors, such as user socio-demographic attributes, vehicle characteristics, land use type of the trip origin, etc., collected from the Australian carsharing company GoGet are utilized for model estimation. A number of forms of parametric hazard-based models are tested to determine the best fit to the data, which was found to be the Weibull distribution in proportional structure. The estimation results of the coefficients of the parameters suggest that users who do not own a private vehicle, frequent users, elderly users and users with a restriction in the ability to drive tend to consider closer vehicle pods. These findings can provide a starting point for carsharing organisations to optimize their pod locations and types of cars available at different pods to maximize usage.

Keywords: Carsharing, Hazard Based Modelling, Choice Set Formation, Vehicle Selection

1. INTRODUCTION

In recent times, transportation planning has focussed on the concept of sustainability of transport systems. The goal of a majority of planning is ensuring a liveable community for the current generation whilst considering the impact on future generations. A number of transportation authorities have recognised the impact that private car ownership has significant costs associated to individuals and departments of transport in relation to purchase and maintenance costs as well as the provision of infrastructure*(1)*. In addition, increased private vehicle use has resulted in traffic congestion. Some of the repercussions due to traffic congestion include; excessive delay, increased fuel consumption, greater road infrastructure costs and higher levels of emissions reducing air quality which has resulted in significant economic and social costs *(2-4)*. In order to mitigate congestion, planners have advocated the use and emphasised the development of public transit, carpooling, walking and cycling. Since the 1980s, as a complement to these approaches, carsharing schemes have become a crucial element of sustainable transport systems within the modern urban cityscape.

Studies have shown that that carsharing schemes have the potential to reduce congestion, provide a more equitable access to private transport and limit the environmental impact of emissions release *(1)*. These advantages have resulted in an increasing presence of carsharing programs as a mode of transport in planning for a sustainable transportation system and community within global urban environments. As carsharing has been successfully adopted by users, it is thus critical to understand what factors affect the demand for carsharing to further their usage. Demand is dependent on trip attributes such as: trip purpose, duration of the trip, time of day and week and also the vehicle selected out of the available choices. Jorge and Correia *(5)*present a comprehensive literature review regarding demand modelling approaches for carsharing programs. The paper highlights that demand estimation is difficult due to the interdependency of availability of vehicles and the number of trips. Furthermore, there has been limited research into understanding and characterising the supply within modelling frameworks. In order to evaluate carsharing programmes effectively, the demand for and the supply provided must be accurately figured. This study aims to bridge this knowledge gap by investigating users' vehicle selection behaviour which is constrained by the supply of vehicles at each carsharing facility.

Users' vehicle selection is a significant factor in determining the most efficient allocation of resources within a carsharing programme. By understanding vehicle selection, programmes can minimise unutilised vehicles within fleet. Thus this study attempts to answer two questions; "How far are users' willing to travel to make use of a carsharing facility/vehicle?" and "What factors influence users' selection of vehicles and are there any patterns or trends associated within these factors?" Users' cognitive capacity for screening and filtering alternatives from a choice set based on a critical or influential factors is an essential component of vehicle selection behaviour as the choice set is extremely large*(6)*. The main factor affecting the choice set of vehicles is considered to be the "distance to the carsharing depot (vehicle pod)" which is formulated using a hazard based approach. The questions of the study are then investigated using a Spatial Hazard Based Model (SHBM), where "distance to the carsharing depot" is considered as a continuous non-negative random variable analogous to the duration of conventional HBMs. A number of parametric forms of HBM are tested to determine the best fit to the data set. The two major contributions of this study are1) introducing an analytical modelling structure for modelling demand for carsharing with a focus on vehicle selection and 2) application of a choice set formation technique that has been previous applied to a housing search problem *(6)*.

The next sections of the paper are structured as follows. Initially a literature review is presented about recent studies conducted within carsharing demand modelling and the application of HBM to problems within the field. The data sets used for the study are then explained highlighting some descriptive statistics. Then the modelling framework and analysis methodology are presented followed by results of the parameter estimation and conclusions of the research.

2. LITERATURE REVIEW

The history and development of carsharing programmes provide the motivation behind this investigation. Carsharing is a form of renting car where people rent cars for short periods, commonly by hour with more flexibility in booking a card and less paper work involved. These programmes have multiple “pick up and drop off” locations allowing users to access vehicles at an origin and return the vehicle at the destination, paying hourly and kilometric fees, thus reducing the need for a personal vehicle *(4, 7, 8)*. In terms of transport planning, carsharing programmes are a travel demand optimisation strategy to reduce the vehicle travelling within the network by providing a user the benefits of a private vehicle when it is absolutely necessary. Martin et al. *(9)* studies the impact of carsharing on household vehicle holdings in North America and presented that the average number of vehicles per household dropped from 0.47 to 0.24. Furthermore, the analysis suggests that carsharing has removed 90,000 to 130,000 vehicles from the road at an aggregate level. A vast amount of literature has highlighted the advantages of carsharing programmes *(1, 5, 8, 10, 11)*. For further information about carsharing, Shaheen and Cohen *(11)* provide the latest overview of the state of practice of carsharing and its impact on transportation systems. The main advantages can be summarised as follows;

- Reduction in costs in relation to travel: Carsharing provides an option to reduce the fixed costs associated with car ownership, such as, insurance, registration and service costs *(12)*. Thus at an individual user level it can encourage saving and more efficiently allocating income.
- Improved accessibility to private vehicle usage: Socially sustainable transport systems can be achieved as lower income earners can now potentially access private vehicles whereas without carsharing schemes it would not be financially feasible *(13)*
- Alleviation in traffic congestion: Carsharing reduces the need for private vehicle ownership and has the ability to reduce the number of vehicles traversing the network. In addition, due to the cost structure of these programmes, trips made by users are planned in advance reducing the number and duration of trips of users, thus reducing total vehicle kilometres travelled.
- Reduction in release of emissions limiting the environmental impact of private vehicle usage
- Improved parking conditions: A reduction in car ownership will reduce the demand for off and on street parking of private vehicles.

Literature suggests the need for carsharing as a mode of transport within the urban environment*(5, 10, 11, 14)*. However the efficient and effective implementation of the programmes is essential to the future success of carsharing and as a result a number of studies *(4, 14-17)*have been conducted to understand what factors determine the demand and supply for carsharing.

User behaviour modelling is one of the branches of carsharing that has recently attracted some attention. Most studies have adopted regression models, stated-preference surveys and data mining techniques to study the characteristics of users of carsharing programmes *(4, 10, 18)*. Catalano et al. *(4)*conducts a stated preference survey within the city of Palermo in Italy. Using the data from 500 respondents the study develops a random utility model and tests future carsharing policy scenarios highlighting the volatility in the impact of planning in carsharing utilisation. Stillwater et al. *(10)*, on the other hand, completed a GIS-based multivariate regression analysis to understand the impact of the built environment and demographic factors of users’ on carsharing demand. Sixteen months of usage data from a carsharing operator in the U.S were used to conduct the analysis which suggested that for one vehicle households, the availability of light rail facilities and age had a positive relationship with demand for carsharing. Morency et al. *(18)* furthered this stream of research by establishing the typology of carsharing users using the carsharing transaction data provided by Communauto, a carsharing company in Montreal, Quebec. The authors used data mining techniques to categorize members based on their temporal units that represented their behaviours. The results indicate a greater proportion of low frequency users (on average 0.4 uses of the program per week) and as a result lower distances travelled (14.3km per week) which is consistent with the aims of the short-term rental principle of carsharing. More recently De Lorimier and El-Geneidy *(19)*developed a multilevel regression model to determine the factors that affect vehicle usage and used a logistics regression analysis to analyse the carsharing vehicle availability. The data used in this research also came from Communauto, a carsharing company in Canada. The results showed that the size of a carsharing station was a key factor to vehicle usage. Morency et al. *(17)* continued investigating the data of Communauto and proposed a two-stage approach to estimate the frequency of usage by an active member. The first stage involved the development of a binary probit model to understand the probability that a member will be active and the second stage was a random utility based model which estimated the probability that a customer will use carsharing mode than once per month given they are an active member. The results indicated that recent usage had a positive relationship with the likelihood of future use; however it is heavily dependent on demographic factors such as age and language spoken at home.

Other than using regression methods or stated preference techniques, Ciari et al. *(14)*presented an activity-based micro simulation model to study carsharing demand. The authors enhanced MATSim, open source simulation software to provide five mode choices for users, i.e. car, public transport, bicycle, walk, and carsharing. The model was applied to the urban area of Zurich, Switzerland. The authors concluded that the results matched the observed demand pattern. However, the features of the carsharing model in this case had no limits on vehicle stocks, which is an unrealistic assumption. Later in 2013, Schaefers *(20)* explored carsharing usage motives using a qualitative means-end chain analysis. They collected data through a series of laddering interviews attended by members of a carsharing company in U.S. The interviews followed the hierarchical structure of means-end chain method. Then, the authors created a hierarchical value map (HVM) to describe the relationships between carsharing attributes and users' core values. They concluded that there are four motivational patterns in carsharing context, namely value-seeking, convenience, lifestyle, and environmental. The results could be used for making carsharing development strategies. Furthermore, Dixit and Rashidi *(21)* investigated safety aspects and factors of carsharing users using a quasi induced exposure method. The study uses data from the GoGet carsharing users' in Sydney, Australia and devised a method to characterise carsharing users involved in crashes. The results of the study indicate that infrequent carsharing users, own at least one vehicle, are more experience and have chosen greater insurance excesses when booking are less likely to be involved in a crash.

Although these studies provide a thorough understanding of some of the significant factors that affect carsharing user behaviour, to the authors' best of knowledge, the process of vehicle selection by a user has yet to be fully investigated.

Recently, there have also been some studies utilised hazard based modelling techniques to explore users' behaviour within carsharing programmes *(17, 22)*. Habib et al. *(22)* further utilised the data provided by Communauto to present a discrete time hazard model to estimate users' decision on continuing membership; a binary probit model to investigate users' decision on using the service; and a zero inflated dynamic ordered probability model to model users' monthly usage. The study modelled the decision-making process over time to consider the history dependence of users' behaviour. Costain et al. *(16)* also applied hazard model to investigate user's demand. They first used descriptive analysis to investigate key attributes that influence the overall patterns of users' behaviour. Then, they developed several econometric models, including binary logit model, hazard model, negative binomial model, multivariate regression model, and multinomial logit model. These models were used to analyse a number of factors including; users' attitude towards environment and safety, frequency of usage and vehicle type choice. Among these models, the hazard model was employed to model membership duration. The results revealed that higher monthly rate and less perceived saving would shorten membership duration. The authors also concluded that carsharing members were more environmentally conscious, and were willing to pay for carbon offsetting. Most members used carsharing service during off-peak period or on weekends for short-distance trips. The methodology behind these studies provided guidance into developing a model which focuses on vehicle selection as this is a discrete decision made based on a set of given alternatives.

The behaviour of vehicle selection is a discrete choice for an individual when utilising carsharing. A user needs to select a single vehicle out of a choice set and how we define the choice set is an important consideration and a key focus of the study. Literature has shown that choice set formation has an impact on the parameter estimation of behavioural choice models *(6, 23, 24)*. Rashidi and

Mohammadian *(6)* presented a detailed review of approaches to choice set formation which has historically been defined into two approaches; random selection out of the universal choice set and consideration of the entire universal choice set, both containing weaknesses in developing accurate behavioural models. In addition, the most critical element within choice set formation is developing an appropriate filtering/screening method *(6, 25)*. The decision process can be completed in two phases; initially determining the probability that an alternative is selected within the choice set and then using this filtered choice set the alternative with the greatest utility can be identified using a secondary choice model *(26)*. This study attempts to achieve the first step of this process through the use of a Spatial Hazard Based Model (SHBM), a relatively new technique that has had a few applications within literature *(6, 27, 28)*. The model considers “distance to the carsharing depot” as a continuous non-negative random variable analogous to the duration of a HBM to construct the choice set of the vehicle selection decision.

3. DATA COLLECTION AND PREPARATION

The carsharing trip data is acquired from GoGet, an Australian carsharing company founded in 2003. 23642 trips are recorded from January 1^{st}, 2012 to June 9^{th}, 2012 to reveal users’ behaviour of selecting carsharing vehicles. These trips are collected by 208 GoGet cars with GPS devices in Sydney, Australia. The data contains a wide range of user-related, vehicle-related, and trip-related variables as shown in Table 1.

TABLE 1 Summary of the Variables Used in the Models

	ID	Variable	Mean	Std. Dev.	Definition
User	1	user_age (year)	38.86	10.63	Age of user
	2	user_dl_year_hold (year)	14.45	8.69	Driving license year held by user
	3	user_car_ownership	0.22	0.42	Binary variable: =1, user owns a car; =0, otherwise
	4	user_how_often_use_ the_car	0.43	0.49	Binary variable: =1, user often uses the car; =0, user rarely uses the car
	5	user_main_way_to_ work	0.58	0.49	Binary variable: =1, user uses public transit to work; =0, otherwise
	6	user_how_many_kms _previous_year (km)	4922.77	22706.55	Distance user traveled by car in the previous year
	7	user_live_near_ dedicated_parking	0.67	0.47	Binary variable: =1, user lives near dedicated parking pod; =0, otherwise
	8	dl_country	0.81	0.39	Binary variable: =1, user's driving license country is Australia; =0, otherwise
	9	dl_lefthand	0.84	0.37	Binary variable: =1, user is from a lefthand driving country; =0, otherwise
	10	user_landuse	0.65	0.48	Binary variable: =1, user's origin landuse type is residential; =0, otherwise
	11	booking_method	0.13	0.34	Binary variable: =1, user uses mobile phone to book a car; =0, otherwise
	12	plan_binary	0.35	0.48	Binary variable: =1, user owns a frequent usage membership plan; =0, otherwise
	13	ins_drug_offences	0.99	0.11	Binary variable: =1, user has not committed drug offences before; =0,
	14	ins_dangerous_drivin g	0.98	0.14	Binary variable: =1, user has not committed dangerous driving before; =0,
	15	ins_traffic_offences	0.68	0.47	Binary variable: =1, user has not committed traffic offences before; =0,
	16	ins_culpable_driving	0.99	0.12	Binary variable: =1, user has not committed culpable driving before; =0,
	17	ins_criminal_charges _ pending	0.99	0.09	Binary variable: =1, user does not have criminal charges pending; =0, otherwise
	18	ins_accident_last_ten _years	0.80	0.40	Binary variable: =1, user does not have traffic accident in the last ten years; =0,
	19	ins_impairment	0.98	0.13	Binary variable: =1, user does not have an impairment; =0, otherwise
Vehicle	20	car_age (year)	3.16	0.74	Age of GoGet car
	21	car_manufacture	0.04	0.20	Binary variable: =1, the manufacturer of the car is Alfa Romeo, =0, otherwise
	22	car_body_type	0.07	0.25	Binary variable: =1, the car is hatchback; =0, otherwise
Trip	23	trip_travel_time	0.22	0.50	Total travel time of each trip
	24	i	0.42	3.52	Number of times user has used a vehicle

Among user-level variables, variables ID 1 to 16 are directly extracted from user information provided by GoGet. The 10th to the 16th user-related attributes reflect whether the user is more aggressive when driving. The 18th attribute is the right-hand/left-hand driving custom, which is not given in the basic user information. To determine whether the user is from countries with right-hand or left-hand traffic, we refer to Wikipedia match user's driving license country to the corresponding driving custom. In addition, user land use binary attribute is obtained by combining the user origin location information and the land use GIS map of the city. GoGet Company provides the coordinates of user origin locations. The land use map comes from the NSW Department of Premier and Cabinet, Office of Environment and Heritage*(29)*, which divides Sydney into 8 zoning categories: residential, commercial and industrial, services, reaction and culture, intensive animal husbandry, intensive horticulture, residual native cover and road. All user origin locations are mapped onto the land use map using ArcGIS, and each piece of user location record is associated with the corresponding Transportation Analysis Zone (TAZ) id and tagged with the corresponding land use type.

Vehicle-related variables contain the age, manufacturer, and body type of the vehicles. The 208 GoGet vehicles being studied are manufactured by 4 manufacturers, i.e., Toyota, Hyundai, Alfa Romeo, and Suzuki. Among them, the rate paid for using Alfa Romeo vehicles is higher than the rate of using the rest of the vehicles. Therefore, a car manufacturer variable is generated as a binary variable that when it equals to 1, the manufacturer is Alfa Romeo; when it equals to 0, the manufacturer is one of the rest companies. Finally, trip-level attributes consist of trip travel time and the number of times that each user selects each car, denoted by i. The number of times a user selects a specific vehicle was determined using a program coded in JAVA that compared the booking time and reservation time of each user given the set of available vehicles. Trip travel time is obtained by subtracting the recorded trip starting time from the recorded trip ending time.

Furthermore, it is also important to note, a maximum catchment of two kilometres that an individual would consider in travelling to a vehicle pod is considered in this study. This assumption has been made to complete an initial screening of the data to determine reasonable and feasible travel distances between the origin and the vehicle pod as walking is the dominant access/egress mode to GoGet vehicles. Figure 1 presents a distribution of the observed GoGet usage data considering the two-kilometre radius. The figure shows a high proportion of users are within one kilometre of the vehicle pod (62% of the total data set) with the usage decreasing as the distance increases. In addition, the total catchment considers 80% of the data set. Accordingly the use of the two-kilometre catchment was deemed as valid initial screening criteria to obtain feasible modelling.

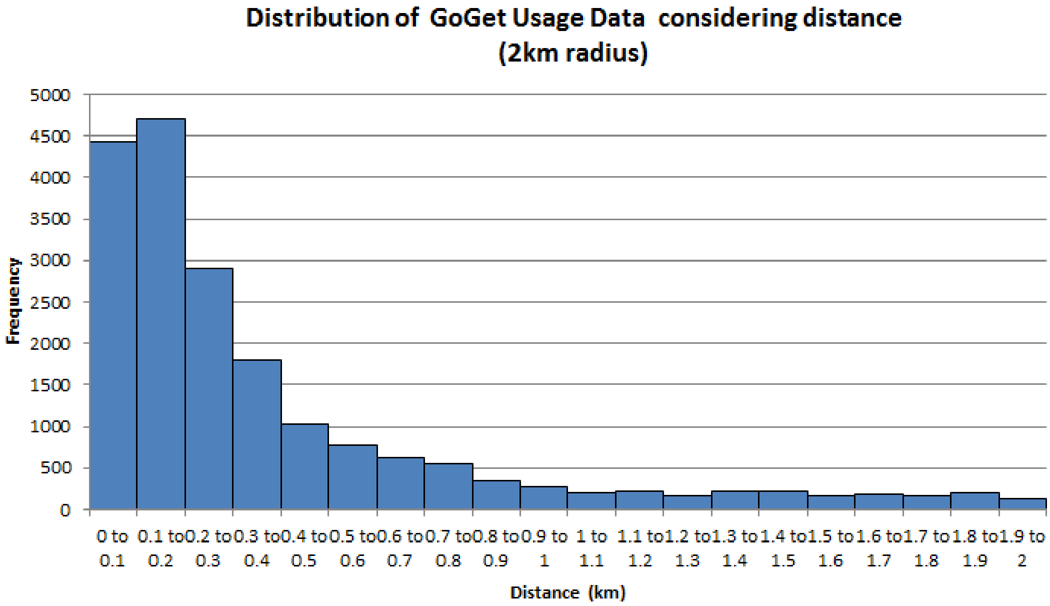

FIGURE 1 Distribution of GoGet Usage Data Considering A 2km Radius

4. MODEL FORMULATION AND METHODOLOGY

In this section, the parametric hazard based models used in this study are introduced, followed by presenting the constraints related to the problem, and the criteria of selecting among a set of parametric models.

In the area of survival analysis, there are two categories of parametric models, namely the accelerated failure time model (AFT model) and the proportional hazards model (PH model). The AFT model considers a linear relationship between the log of survival time and covariates while a proportional hazard model assumes that absolute differences in covariates imply proportionate differences in the hazard rate at a specific time. There are different probability distributions, such as exponential distribution, Weibull distribution, log-logistic distribution, etc., employed to formulate parametric hazard models. Among these distributions, Weibull function has been most frequently used in the studies of duration modelling*(6, 30)*since Cox*(31)*first proposed the Weibull baseline hazard model in 1959. It presents a monotonically increasing or decreasing hazard function. However, it should be noted that the hazard function might not be monotonic in some cases, which needs us to test non-monotonic distributions for the parametric hazard models.

Although PH models are more widely used in the survival analysis, we also examine AFT models in this research because the regression parameter estimates of AFT models are more robust to omitted covariates, and therefore, less influenced by the selected probability distribution. Furthermore, we test both monotonic functions (i.e. Weibull and exponential distributions) and non-monotonic functions (i.e. log-logistic and lognormal distributions) to figure out the most fitted parametric hazard model.

As discussed in the literature review, duration and distance in all the formulations presented in this section are interchangeable without losing the generality. In the hazard formulation, the length of a spell for a subject, e.g. a user and a vehicle, is represented by a continuous random variable T with a cumulative density function (CDF), $F(t)$, and probability density function (PDF), $f(t)$. The probability of failing sometime before time *t*, $F(t)$ can be written as:

$$\Pr(T \leq t) = F(t) \tag{1}$$

where *t* denotes the elapsed time since entry to the state at time 0.

Then, the survival function can be written as:

$$\Pr(T > t) = 1 - F(t) = S(t) \tag{2}$$

As the slope of CDF, PDF is:

$$f(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \leq T \leq t + \Delta t)}{\Delta t} = \frac{\partial F(t)}{\partial t} = -\frac{\partial S(t)}{\partial t} \tag{3}$$

where Δt is a very small time interval.

The hazard rate is defined as the probability of failure in the interval $(t, t + \Delta t]$given that it has survived until time t. Let $\theta(t)$denote the hazard rate, we have:

$$\theta(t) = \frac{f(t)}{1 - F(t)} = -\frac{f(t)}{S(t)} = \frac{\partial\{-ln[S(t)]\}}{\partial t} \tag{4}$$

Using Equation (4), the survival function is:

$$S(t) = \exp\left[-\int_0^t \theta(u) du\right] \tag{5}$$

As mentioned before, we use AFT models to conduct parameter estimates. The AFT parametric models are formulated as:

$$ln(t_j) = x_j\beta_x + z_j \tag{6}$$

where x_j represents explanatory variables, $\beta_x$denotes the vector of coefficients, and z_j is the error with density $f(.)$, which determines the regression model. By letting $f(.)$ be the extreme-value density, the Weibull and exponential parametric hazard models are obtained. By setting $f(.)$ equal following a logistic distribution, the log-logistic hazard model is formulated. Similarly, by setting $f(.)$ to a normal distribution, we obtain a lognormal regression model.

For AFT model, $\exp(-x_j\beta_x)$ is defined as the acceleration parameter. If $\exp(-x_j\beta_x)$ equals to 1, time passes at its normal rate. If $\exp(-x_j\beta_x)$ is larger than 1, time is accelerated, that is, the failure is expected to occur sooner. If $\exp(-x_j\beta_x)$ is smaller than 1, then time is decelerated, which means the failure might occur later.

Using Equation (5), (6), and the density functions, we rewrite the survival functions of these four distributions as follows:

Exponential survival function: $$S(t) = \exp\left[-\exp(-x_j\beta_x)t_j\right] \tag{7}$$

Weibull survival function: $$S(t) = \exp\left[-\exp(-px_j\beta_x)t_j^p\right] \tag{8}$$

Log-logistic survival function: $$S(t) = \left(1 + \left[\exp(-x_j\beta_x)t_j\right]^{1/\gamma}\right)^{-1} \tag{9}$$

Lognormal survival function: $$S(t) = 1 - \phi\left(\frac{\ln(t_j) - x_j\beta_x}{\sigma}\right) \tag{10}$$

where p is the shape parameter of the Weibull distribution, γ is the shape parameter of the log-logistic distribution, σ is the shape parameter of the lognormal distribution, x_j represents the vector of covariates, and β_x is the vector of coefficients.

It should be noted that in the AFT models, the effect of covariates is facilitated by incorporating a negative sign for the parameters in formulation. In other words, if the coefficient (β_x) of the covariate is estimated to have a negative value, the expected time to failure increases and the probability of failure decreases. In the context of the distance selection of this paper, this means that an individual with a larger positive value of the covariate tends to choose vehicles with shorter distances between the origin and vehicle pod. Conversely, if β_x of a covariate is positive, an increase in the value of the covariate decreases the expected time to failure and therefore increases the probability of failure. This means that an individual with a higher value of this type of covariate tends to increase the distance between the origin and vehicle pod.

Since we formulate four different hazard models in this research, it is necessary to compare the goodness of fit with all four models. The Akaike Information Criterion (AIC) and Bayesian Information Criteria (BIC) are employed to select the best-fitted hazard model. Both of them are criteria for model selection among a finite set of parametric models. The model with lower AIC and BIC value is considered to be better fitted. The formulations of AIC and BIC are:

$$AIC = -2 \times \ln(likelihood) + 2 \times k \tag{11}$$

$$BIC = -2 \times \ln(likelihood) + \ln(N) \times k \tag{12}$$

where k denotes the number of the parameters estimated, and N is the number of observations.

5. RESULTS AND ANALYSIS

The hazard based model estimation was conducted using the statistical analysis software package STATA12 for all four models considered for the study. Table 2 shows the results of parameter estimation. The values of variable coefficients are very close in all four models. The signs of those coefficients are also consistent in all models. The p parameter of Weibull hazard model is greater than 1, which determines that the shape of the model is monotonically increasing. The γ parameter of log-logistic distribution and the σ parameter of lognormal distribution are smaller than 1 indicating that the shapes of those two models are also monotonically increasing. Figure 2 presents the patterns of hazard values for the four hazard function types. Except the exponential model, the three other models have the same patterns that the hazard value increases as the distance between the user and the vehicle increases. This is in line with the common sense that people prefer to select vehicles that are nearer to their origin locations.

TABLE 2 Parameter Estimation Results

		Weibull		Exponential		Loglogistic		Lognormal	
	Parameter	Estimate	z value	Estimate	z value	Estimate	z value	Estimate	z value
User	user_age	-0.00051	-8.97	-0.000757	-4.92	-0.000449	-6	-0.000274	-3.08
User	user_dl_year_hold	0.0001256	1.85	0.0000365	0.2	-0.000187	-2.08	-0.000493	-4.6
User	user_car_ownership	-0.009502	-8.34	-0.022392	-7.23	-0.013791	-9.09	-0.019062	-10.74
User	user_how_often_use_the_car	0.0162167	16.98	0.0373839	14.46	0.0262326	20.83	0.0393404	26.36
User	user_main_way_to_work	-0.00197	-2.14	-0.004364	-1.75	-0.003859	-3.17	0.0008672	0.6
User	user_how_many_kms_previous_year	4.65E-07	10.32	7.67E-07	6.32	5.60E-07	9.72	2.28E-07	7.47
User	user_live_near_dedicated_parking	0.0082488	8.46	0.0085782	3.24	0.0163726	12.7	0.0101486	6.65
User	dl_country	-0.013815	-5.26	-0.015169	-2.12	-0.021788	-6.28	-0.01042	-2.52
User	dl_lefthand	0.0196038	7.01	0.0263786	3.47	0.0304322	8.24	0.0207229	4.71
User	user_landuse	-0.009483	-9.79	-0.019233	-7.32	-0.011136	-8.7	-0.007726	-5.09
User	booking_method	-0.002589	-1.92	-0.008056	-2.2	-0.000734	-0.41	-0.000642	-0.3
User	plan_binary	-0.00571	-5.71	-0.015311	-5.64	-0.003066	-2.31	-0.004479	-2.85
User	ins_drug_offences	-0.047043	-4.78	-0.157879	-5.95	-0.094632	-7.2	-0.103237	-6.58
User	ins_dangerous_driving	0.0252194	4.43	0.0725084	4.68	0.0517034	6.56	0.0445719	4.87
User	ins_traffic_offences	0.0132251	13.32	0.0277416	10.31	0.016577	12.58	0.0122849	7.87
User	ins_culpable_driving	0.0100419	1.11	0.0591836	2.42	0.0107754	0.87	0.0208308	1.42
User	ins_criminal_charges_pending	0.022081	2.76	0.0681017	3.14	0.0713973	6.85	0.1079063	8.72
User	ins_accident_last_ten_years	-0.02044	-17.05	-0.044311	-13.7	-0.020767	-13.27	-0.023518	-12.65
User	ins_impairment	-0.047551	-10.44	-0.085874	-6.95	-0.08257	-14.22	-0.094441	-13.4
Vehicle	car_age	0.0024852	4.21	0.0026424	1.63	0.0055059	6.97	0.0032347	3.42
Vehicle	car_manufacturer	0.0411433	18.76	0.0419121	7.03	0.0594497	20.53	0.0386941	11.21
Vehicle	car_body_type	-0.008204	-4.64	-0.003931	-0.82	-0.010142	-4.33	-0.000796	-0.29
Trip	trip_travel_time	0.0053212	5.56	0.0173492	6.67	0.0101375	8.32	0.0127223	8.88
Trip	I (Number of time user has used a vehicle)	-0.014887	-170.1	-0.028872	-120.9	-0.076393	-233.7	-0.055436	-281.6
	_cons	0.3986554	68.14	0.3490078	21.93	0.2481781	32.18	0.1515855	16.56
	Weibull_ln_p	1.000219	980.2	-	-	-	-	-	-
	Weibull_p	2.718877	-	-	-	-	-	-	-
	Loglogistic_ln_gamma	-	-	-	-	-1.231951	-1201	-	-
	Loglogistic_gamma	-	-	-	-	0.2917228	-	-	-
	Lognormal_ln_sigma	-	-	-	-	-	-	-0.531353	-624.1
	Lognormal_sigma	-	-	-	-	-	-	0.5878091	-

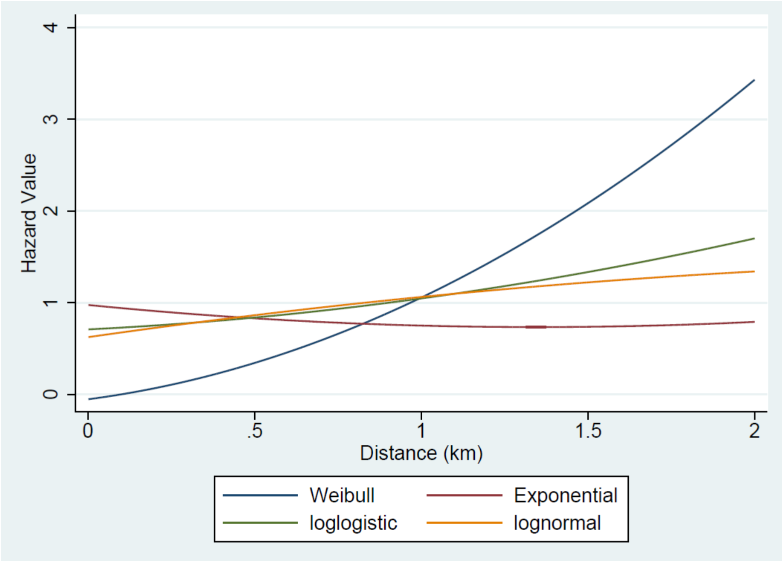

FIGURE 2 Hazard Patterns for Weibull, Exponential, Log-logistic, and Lognormal Functions

As discussed in the Model Formulation and Methodology section, AIC and BIC are employed to select the best-fitted hazard model. With Equation (11) (12), we obtain Table 3 showing the results of AIC and BIC of all four models. The AIC and BIC values for the model with the Weibull distribution are 952934.7 and 953234.1, respectively, which are the smallest among all the models. It can be concluded that the Weibull distribution provides the best hazard model for the vehicle selection data among the four distributions.

TABLE 3 Calculation Results of AIC and BIC of Four Models

	Weibull	Exponential	Log-logistic	Lognormal
Log likelihood value with only constant	-483999.9	-809987.7	-598379.6	-667916.9
Log likelihood value at convergence	-476441.4	-805768.8	-568078.4	-629825.6
Number of observations	739973	739973	739973	739973
Number of parameters	26	25	26	26
AIC	952934.7	1611588	1136209	1259703
BIC	953234.1	1611875	1136508	1260003

The parameter estimation for the Weibull distribution shows that all coefficients of the covariates are statistically significant at a 5% confidence level, as presented in Table 2. It is clear that car ownership has a negative sign suggesting that users' who are located closer to the pod own a car of their own and will be using the car sharing service more often. In other words, people not owning a car but willing to use the service of carsharing to travel a greater distance to the vehicle pod. However people who own a car have the choice to not use a carsharing vehicle if the distance to the vehicle pod is too great. In addition the users' age has a negative coefficient suggesting the older a user the closer his or her origin is to the vehicle pod, which is a reasonable observation as the act of travel becomes a more difficult task for elderly people. The covariate "dl_country_binary", also has a negative coefficient suggesting that users with Australian drivers' licenses originate closer to the vehicle pod than users' that don't have a domestic license. Again this is a rational outcome as there would be a proportion of users' who are travelling or temporarily residing within the country and accordingly their knowledge of pod locations and knowledge of access to Go Get carsharing program could be limited. The car body type variable also has a negative coefficient estimate suggesting that users' selecting hatchback vehicles are less willing to travel longer distances to the vehicle pods. Hatchbacks are the most common within the GoGet vehicle fleet resulting in the users' travelling shorter distances to utilise these car body types. Furthermore, it is interesting to note that the other covariates with the greatest negative coefficients are related to people having drug offences, accidents within the last 10 years or some sort of impairment. These characteristics indicate reluctance to drive and own a car which shows a higher likelihood of living as close as possible to carsharing facilities to cater for the occasional use of a car.

Observing the positive coefficients, the greatest positive coefficient is related to the binary car manufacturer variable which indicates that people who used an Alfa Romeo as the GoGet vehicle had the tendency to commute further to the vehicle pod. This finding can be explained by the fact that these luxury cars are not necessarily close to the user but because of the extra utility the user is expecting from the vehicle accepts the disutility of distance. Nonetheless, further investigation needs to be completed to understand the impact of all the different car brands and types and how that impacts the distance to travel to the vehicle pod as it is expected different brands would have different associated coefficients. In addition increased usage of cars (user_how_often_use_the_car_binary) also increases the distance between the origin and the vehicle pod. There are a few explanations for this result; initially owners of private vehicles are less likely to use carsharing facilities so as a result there is no utility for people to be within the vicinity of the carsharing service. Within Australia, carsharing vehicles have priority parking spaces throughout built up CBD areas where parking is a premium and accordingly users' may find it more convenient to use the carsharing facility ahead of using their own private vehicle. Another interesting observation is that the covariates which describe the drivers who have been suspended or lost their licence, due to the increased number of driving offences, all show positive coefficients. These covariates observe events which are not expected by users'. If the users' are dependent on travelling with their own private vehicle they would not normally consider carsharing as a mode. Thus the origin of travel will be further from the vehicle pod in a similar fashion to users' that use their car on a regular basis.

Overall these results suggest that frequent users of GoGet, elderly users and users with a restriction in the ability to drive tend to originate closer to the vehicle pod. On the other hand, users' who travel by car more frequently, select luxury vehicles (Alfa Romeo) as a carsharing vehicle, and have had driving offences originate further away from the vehicle pod. Based on these trends carsharing programs can optimise their vehicle pod locations to centre on catchments containing these user classes to maximise usage of the system as well as enhance the overall popularity of the scheme.

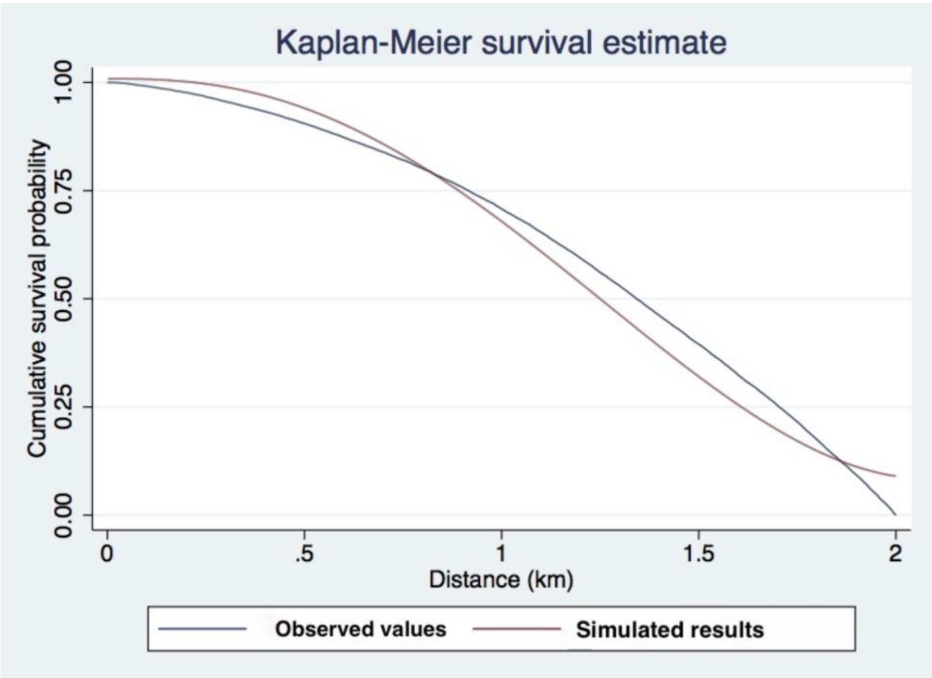

FIGURE 3 Comparisons between the Observed and the Simulated Results

The hazard model with Weibull function which provided the best fit to the data was further examined. The CDF is employed to randomly simulate the distance between a GoGet user and his or her selected vehicle.Figure3 shows the comparison between the simulated results for the estimated hazard model with Weibull distribution and the observed values. It can be concluded from the figure that the general patterns of the observed result and the simulated result are relatively similar. There is a small gap that can be seen between these two results.

6. CONCLUSION AND FUTURE DIRECTIONS

This study presented a behavioural model to gain a greater understanding of users' selecting vehicles within a carsharing program. The study attempted to answer two questions; "How far are users' willing to travel to make use of a carsharing facility/vehicle?" and "What factors influence users' selection of vehicles and are there any patterns or trends associated within these factors?" An answer to these two questions enable the researcher to model the choice set formation behaviour as a probabilistic process which is a function of distance the identified covariates.

As the vehicle selection process is complicated a choice set formation methodology using a spatial HBM was developed and validated using a rich data set from the Australian carsharing company GoGet. The SHBM considered "distance to the carsharing depot (vehicle pod)" as a continuous non-negative random variable analogous to the duration of conventional HBMs. The results from the modelling contain a number of negatively and positively correlated covariates which can provide trends and patterns that could potentially be used to guide policy of carsharing programmes. The number of times a user selects a GoGet vehicle, user age, car ownership, having an Australian drivers' license and variables indicating the presence of drug offences, impairment to driving and accidents all had negative coefficient estimates imply the tendency toward shorter distance as the variables increases. This indicates that the impairment and immobility of users affect the proximity while local users' awareness of the availability of carsharing results in shorter distances to the vehicle pod. The car manufacturer, frequency of usage of cars and the variables considering the instances of traffic offences all had positive coefficient estimates. This may suggest luxury usage of the carsharing programmes and users' who regularly travel in private vehicles tend to originate further from the vehicle pods as these tend to be infrequent trips. The results of the modelling suggest that users that frequent users of GoGet, elderly users and users with a restriction in the ability to drive tend to originate closer to the vehicle pod. Based on these trends carsharing programmes can optimise their vehicle pod locations to centre on catchments containing these user classes to maximise usage of the system as well as enhance the overall popularity of the scheme.

Further extensions of the current study include the investigating the importance of other key variables other than the "distance to the carsharing depot (vehicle pod)" that impact the vehicle selection process within a carsharing scheme such as; trip purpose and trip destination. In addition the impact of the types of vehicles (hatchback/4WD/minivan) and the manufacturer of the vehicles could also be further refined as individual covariates to see if an impact is observed.

7. ACKNOWLEDGEMENTS

The authors would like to thank Mr. Bruce Jeffreys and Ms. Rachel Moore from GoGet for providing data. We would also like to thank the Australian Research Council for their support under Linkage Grant # LP130100983.

8. REFERENCES

1. Duncan, M., *The cost saving potential of carsharing in a US context.* Transportation, 2011. 38(2): p. 363-382.

2. Wijayaratna, K., *Reliability of the South-West Sydney Transport Corridor*, in *School of Civil and Environmental Engineering*2010, University of New South Wales.

3. Banister, D., *Unsustainable transport: city transport in the new century*2005: Taylor & Francis.

4. Catalano, M., B. Lo Casto, and M. Migliore, *Car sharing demand estimation and urban transport demand modelling using stated preference techniques*. 2008.

5. Jorge, D. and G. Correia, *Carsharing systems demand estimation and defined operations: a literature review*. EJTIR, 2013. 13(3): p. 201-220.

6. Rashidi, T.H. and A.K. Mohammadian, *Behavioral Housing Search Choice Set Formation: A Spatial Hazard-based Screening Model*. International Regional Science Review, 2012: p. 0160017612461356.

7. Bonsall, P., A. Spencer, and W.-S. Tang, *Ridesharing in Great Britain: Performance and impact of the YORKSHARE Schemes*. Transportation Research Part A: General, 1983. 17(3): p. 169-181.

8. Shaheen, S., D. Sperling, and C. Wagner, *Carsharing in Europe and North American: Past, Present, and Future*. 1998.

9. Martin, E., S.A. Shaheen, and J. Lidicker, *Impact of carsharing on household vehicle holdings*. Transportation Research Record: Journal of the Transportation Research Board, 2010. 2143(1): p. 150-158.

10. Stillwater, T., P.L. Mokhtarian, and S.A. Shaheen, *Carsharing and the built environment*. Transportation Research Record: Journal of the Transportation Research Board, 2009. 2110(1): p. 27-34.

11. Shaheen, S.A. and A.P. Cohen, *Carsharing and personal vehicle services: worldwide market developments and emerging trends*. International Journal of Sustainable Transportation, 2013. 7(1): p. 5-34.

12. Katzev, R., *Car Sharing: A New Approach to Urban Transportation Problems*. Analyses of Social Issues and Public Policy, 2003. 3(1): p. 65-86.

13. Wachs, M. and B.D. Taylor, *Can Transportation Strategies Help Meet the Welfare Challenge?* Journal of the American Planning Association, 1998. 64(1): p. 15-19.

14. Ciari, F., N. Schuessler, and K.W. Axhausen, *Estimation of Carsharing Demand Using an Activity-Based Microsimulation Approach: Model Discussion and Some Results*. International Journal of Sustainable Transportation, 2013. 7(1): p. 70-84.

15. Correia, G.H.d.A. and A.P. Antunes, *Optimization approach to depot location and trip selection in one-way carsharing systems*. Transportation Research Part E: Logistics and Transportation Review, 2012. 48(1): p. 233-247.

16. Costain, C., C. Ardron, and K.N. Habib, *Synopsis of users' behaviour of a carsharing program: A case study in Toronto*. Transportation Research Part A: Policy and Practice, 2012. 46(3): p. 421-434.

17. Morency, C., K.M.N. Habib, V. Grasset, and M.T. Islam, *Understanding members' carsharing (activity) persistency by using econometric model*. Journal of advanced Transportation, 2012. 46(1): p. 26-38.

18. Morency, C., M. Trepanier, and B. Agard. *Typology of carsharing members*. in *Transportation Research Board 90th Annual Meeting*. 2011.

19. De Lorimier, A. and A.M. El-Geneidy, *Understanding the factors affecting vehicle usage and availability in carsharing networks: A case study of Communauto carsharing system from Montréal, Canada*. International Journal of Sustainable Transportation, 2013. 7(1): p. 35-51.

20. Schaefers, T., *Exploring carsharing usage motives: A hierarchical means-end chain analysis*. Transportation Research Part A: Policy and Practice, 2013. 47: p. 69-77.

21. Dixit, V. and T.H. Rashidi, *Modelling crash propensity of carshare members*. Accident Analysis & Prevention, 2014. 70: p. 140-147.

22. Habib, K.M.N., C. Morency, M.T. Islam, and V. Grasset, *Modelling users' behaviour of a carsharing program: Application of a joint hazard and zero inflated dynamic ordered probability model.* Transportation research part A: policy and practice, 2012. 46(2): p. 241-254.

23. Ben-Akiva, M.E. and S.R. Lerman, *Discrete choice analysis: theory and application to travel demand*. Vol. 9. 1985: MIT press.

24. Timmermans, H. and R.G. Golledge, *Applications of behavioural research on spatial problems II: preference and choice.* Progress in human geography, 1990. 14(3): p. 311-354.

25. Manski, C.F., *The structure of random utility models.* Theory and decision, 1977. 8(3): p. 229-254.

26. Lerman, S.R., *Recent advances in disaggregate demand modelling.* Publication of: Elsevier Science Publishers BV, 1984.

27. Rogerson, P.A., R.H. Weng, and G. Lin, *The spatial separation of parents and their adult children.* Annals of the association of American Geographers, 1993. 83(4): p. 656-671.

28. Pellegrini, P.A. and J.T. Grant, *Policy coalitions in the US Congress: A spatial duration modeling approach.* Geographical Analysis, 1999. 31(1): p. 45-66.

29. NSW Government Environment and Heritage, *Land Use: New South Wales*, L.A.U.a.I.M. Brokerage, Editor 2014.

30. Yamamoto, T. and R. Kitamura, *An analysis of household vehicle holding durations considering intended holding durations.* Transportation Research Part A: Policy and Practice, 2000. 34(5): p. 339-351.

31. Cox, D.R., *The analysis of exponentially distributed life-times with two types of failure.* Journal of the Royal Statistical Society. Series B (Methodological), 1959: p. 411-421.