AN INTEGRATED APPROACH ANALYZING THE HOUSEHOLD VEHICLE TYPE CHOICE, TRAVELLING DISTANCE, AND HOLDING DURATION BASED ON A COPULA MODEL

Masashi KUWANO
Transport Studies Group, Graduate School of Engineering,
Hiroshima University, Kagamiyama 1-4-1, Higashi-Hiroshima, 739-8527, Japan
E-mail: kuwano@hiroshima-u.ac.jp / Phone & Fax: +81-82-424-7825

Akimasa FUJIWARA
Transport Studies Group, Graduate School for International Development and Cooperation, Hiroshima University, Kagamiyama 1-5-1, Higashi-Hiroshima, 739-8529, Japan; E-mail: afujiw@hiroshima-u.ac.jp / Phone & Fax: +81-82-424-6921

Junyi ZHANG
Transport Studies Group, Graduate School for International Development and Cooperation, Hiroshima University, Kagamiyama 1-5-1, Higashi-Hiroshima, 739-8529, Japan; E-mail: zjy@hiroshima-u.ac.jp / Phone & Fax: +81-82-424-6919

and

Makoto TSUKAI
Transport Studies Group, Graduate School of Engineering,
Hiroshima University, Kagamiyama 1-4-1, Higashi-Hiroshima, 739-8527, Japan
E-mail: mtukai@hiroshima-u.ac.jp / Phone & Fax: +81-82-424-7827

ABSTRACT

The objectives of this study are twofold: 1) to develop an integrated approach analyzing the vehicle type choice, annual travelling distance, and holding duration by simultaneously consider the correlations between the three behaviours, and 2) to demonstrate the changes in CO_2 emissions under different taxation policies. In order to develop the integrated approach, focusing on stages of vehicle usage and vehicle ownership, this study proposes a Copula-based Multivariate Survival (CMS) model of holding duration and annual travelling distance. A Paired Combination Logit (PCL) is applied in order to model the vehicle type choice, and then the estimated PCL model is further incorporated into the CMS model. An empirical analysis was carried out by using a data set collected in the Chugoku region of Japan, 2006. Model estimation results empirically confirmed the effectiveness of the suggested approach. The estimation results also showed that the vehicle holding duration has the negative correlations with expected utilities of vehicle type choice and with vehicle usage. Through the simulation analysis about vehicle-related taxes, it was clarified that to increase fuel tax is the most effective in to reducing the CO_2 emissions, followed by auto tax and weight tax at the vehicle inspection. Moreover, it was observed that the increases in acquisition tax contribute to increase CO_2 emissions.

Keywords: Integrated approach, vehicle type, annual travelling distance, holding duration, copula

1. INTRODUCTION

In most countries, energy consumption and the resulting pollutants from transport sector have been increasing considerably in the past decades. Considering the growing concerns of global warming, the effective measures to control private vehicle use, to enhance efficient use of energy, and to reduce emissions throughout transportation systems, are required. Since the passenger vehicles amount to large percentage in the total exhausted gases, this paper focuses on the household ownership behaviour of passenger vehicles. The household vehicle ownership behaviour consists of the following three stages as “vehicle purchase”, “vehicle usage”, and “duration of vehicle ownership” (De Jong, 2004). Three these stages are dependent over time. Existing studies have been attempted to describe the complicated vehicle ownership behaviour. For example, Bhat and Sen (2006) applied discrete-continuous models to deal with the choice of vehicle type and the usage of vehicle, simultaneously. However, no studies to incorporate the correlation between the stages of vehicle usage and holding duration are accumulated. Recently, in Japan, a rapid increase in fuel prices caused the decrease in vehicle use on the macro level, the shift in vehicles with more fuel-efficient engines, and a decrease in private vehicle ownership. There is no doubt that change in the running cost of private vehicle affect on vehicle type choice, vehicle usage, and its holding duration. On the other hand, in order to reduce total emissions from passenger vehicles, the promotion in the ownership of low emission vehicles, and reducing vehicle use frequency or trip distance become more and more important. Vehicle ownership behaviour is also affected by the household and individual attributes, together with the policy variables such as tax, service levels of transit systems, and land use, etc. So then the model should appropriately consider these variables. In case of the vehicle related taxation change, the policy maker should care for the difference in those taxes are collocate at different stages. For example, those who want to buy a vehicle in Japan, have to pay acquisition tax and consumption tax, and also have to pay annual holding tax (i.e., auto tax) addition to the weight tax at vehicle inspection in every two years during the stage of owning the vehicle. Needless to say, those people have to pay fuel taxes. It is considered that three types of taxes paid at different stages would differently influence on household vehicle ownership behaviour. In order to find the effective tax policies, vehicle ownership behaviour should be integrated in the model system incorporating the interdependency between these behaviours. In other words, vehicle ownership behaviour analysis based on the assumption of the independence across the three stages may lead into the misunderstandings of the behaviour.

The first objective of this study is to develop an integrated approach analyzing vehicle purchase, vehicle usage, and duration of vehicle ownership simultaneously by accommodating the correlations between the three behaviours. The second objective is to demonstrate the effective policies to reduce the environmental emissions from passenger vehicles. In order to develop the integrated approach, firstly, we formulate the vehicle type choice model based on a discrete choice model. And then, in terms of the stages of vehicle usage and duration of vehicle ownership, this study proposes a Copula-based Multivariate Survival (CMS) model of annual travelling distance and holding duration, and finally combines the model with the vehicle type choice model. “Copula” is a function to approximate a stochastic correlation among the random variables under the preliminary marginal distributions (Nelsen, 2006). The copula function is a simple and powerful approach to give a closed-form analytic expression for the joint probability, and flexible to represent the correlated random variables which may follows different marginal distributions. Moreover, the copulas can be applied to the general multivariate models to integrate various types of correlation structures.

This paper is organized as follows. The overall structures of the model system and model formulations are described in Section 2. In Section 3 we will describe the dataset. The model estimations are reported in the Section 4. In Section 5, we conduct simulation analysis about the impact of the tax changes in different stages. Finally, Section 6 shows some conclusions and directions for the further research.

2. MODEL DEVELOPMENT

2.1 Modelling framework

In this study, we develop an integrated model simultaneously including "vehicle purchase", "vehicle usage", and "duration of vehicle ownership". The stage of "vehicle purchase" consists of two behaviours: 1) choice of number of vehicles, and 2) vehicle type. Since the private vehicle market is already saturated in Japan, there is no significant increase in number of vehicles over the past decades. Therefore, this study only models the vehicle type choice in vehicle purchase stage, and in order to describe the behaviour following random utility maximization, a disaggregate choice model is applied.

Concerning to the vehicle usage and the duration of vehicle ownership in day to day horizon, these behaviours can be further decomposed into each of behaviour such as the purpose of vehicle use, destination choice, modal choice, and the departure timing decision. However, each behaviour cannot be modelled without using the detailed data, therefore these models are out of concern in this study. This study focuses on annual travelling distance and holding duration, and applies the survival model to analyze them. Even though the survival model is not a behaviourally-oriented model, it can appropriately describe the probability density of these continuous variables, and flexibly design the model frame. The remaining sub-sections discuss the characteristics of the models at each stage of vehicle ownership.

2.2 Vehicle Type Choice Modelling

Disaggregate choice models are generally used to describe the choice behaviour of vehicle type, where household characteristics, and vehicle attributes are used as explanatory variables in the models. Moreover, it is expected that vehicle use preference significantly influence on the vehicle type choice. As discussed in 2.1, the detailed individual characteristics in vehicle use purpose, frequency and so on are not modelled in this study. Therefore, such unobserved factors in vehicle use are considered as the latent characteristics in our model. Therefore, we introduce annual travelling distance in the last year as a proxy variable of the vehicle use preference.

From the perspective of modelling the vehicle type choice, it has been recognized that the unobserved correlations between the alternatives should be considered, otherwise the biased estimates are obtained (Brownstone and Train, 2000). This study applies the Paired Combination Logit (PCL) model to represent the correlations. PCL model was proposed by Chu (1989) and further studied by Koppleman and Wen (2000) from the theoretical aspects in terms of the model structure, properties, and estimation. The model allows the correlation between any pair of alternatives. The joint partial dependency among the alternatives, are added as similarity coefficients $\sigma_{jj'} \in (0,1)$ where the alternative j and j' are identical if $\sigma_{jj'}$ takes 1, while the alternative j and j' are independent if $\sigma_{jj'}$ takes 0.

The choice probability is described by:

$$P(j)=\frac{\sum_{j\neq j'}\left(1-\sigma_{jj'}\right)\left\{\exp\left(\frac{V_j}{1-\sigma_{jj'}}\right)+\exp\left(\frac{V_{j'}}{1-\sigma_{jj'}}\right)\right\}^{-\sigma_{jj'}}\exp\left(\frac{V_j}{1-\sigma_{jj'}}\right)}{\sum_{q=1}^{n-1}\sum_{r=q+1}^{n}\left(1-\sigma_{qr}\right)\left\{\exp\left(\frac{V_q}{1-\sigma_{qr}}\right)+\exp\left(\frac{V_r}{1-\sigma_{qr}}\right)\right\}^{1-\sigma_{qr}}}$$
$$=\sum_{j\neq j'}P(jj')P(j\,|\,(jj')) \quad (1)$$

where, $\sigma_{jj'}$ is the similarity coefficient between alternative j and j', V_j is the non-stochastic term of the utility for the alternative j, $P(jj')$ is the marginal probability for the alternative pair j and j', and $P(j\,|\,(jj'))$ is the conditional probability of choosing alternative j is given that the alternative pair jj' has been chosen.

2.3 Joint modelling of holding duration and annual travelling distance

Survival analysis has been extensively used in econometrics, biostatistics, medical sciences, and industrial engineering in order to model the p.d.f. of the continuous variable before the concerning event. In the area of transportation, we can find some studies applying it to the analysis of the duration of a vehicle holding terms (Hensher 1985; Gilbert 1992; Yamamoto and Kitamura 2000), the duration of activities (Mannering and Hamed 1990), and the duration between vehicle accidents (Mannering et al. 1994). In this study, the vehicle annual travelling distance (d) and vehicle holding duration (t) are examined using survival models.
The annual travelling distance is usually influenced by a number of factors. In this study, we introduce household attributes, main-user attributes, and vehicle attributes as the covariates in our model. These covariates are also used in the model of vehicle holding duration. In addition, the logsum variable estimated in the vehicle type choice models is introduced as a covariate, since it can be an index of the price and quality of vehicles available in the current market. The coefficient of the vehicle type logsum would be negative, because the higher expected utility in vehicle choice alternatives is, the shorter duration of vehicle holding is.
In 2009 a rapid increase in fuel price caused the decrease of vehicle kilometres travelled and the increase of vehicle replacement into the smaller size. There is no doubt that the change in vehicle running cost affects both on vehicle use and holding duration. In other words, the annual travel distance and the holding duration are dependent. This study attempts to propose a Copula-based Multivariate Survival (CMS) model to capture the interdependence between these two behaviours.
In the next section, we firstly formulate a uni-variate survival model for analyzing the annual travelling distance and vehicle holding duration, and then develop a multivariate survival model based with copula functions.

2.3.1 A Uni-variate Survival Model

In a survival model, time T is considered as a continuous random variable. It measures the duration up to occurring the concerning event. In this study, it is used to represent ownership duration of a vehicle and annual travelling distance. Suppose that T has a continuous probability density function $f(t)$. The distribution function ($F(t)$) gives the probability that the failure time is less than or equal to t:

$$F(t)=\int_0^t h(s)ds=\Pr[T\le t]\,. \tag{2}$$

Then, the hazard function can be written as a function of the distribution function ($F(t)$) and the corresponding density function ($f(t)$) of the random variable t.

$$h(t)=\frac{f(t)}{1-F(t)} \tag{3}$$

Another important function in hazard-based models is the survival function ($S(t)$), which gives the probability of duration t still not to occur the concerning event. It is related to the distribution function as follows:

$$S(t)=\Pr(T\ge t)=1-\Pr(T\le t)=1-F(t)\,. \tag{4}$$

Since $f(t)=-dS(t)/dt$, the hazard can also be written as:

$$h(t)=-\frac{d(\log S(t))}{dt}\,. \tag{5}$$

If the hazard is given, the survival function can be found through:

$$S(t)=\exp(-\int_0^t h(u)du)\,. \tag{6}$$

And then, the density function of t is expressed by:

$$f(t)=h(t)\exp(-\int_0^t h(t)dt)\,. \tag{7}$$

As shown in eq. (2) to (7), $f(t)$, $S(t)$ and $h(t)$ are mathematically identical. If the distribution $f(t)$ is given, then $S(t)$ and $h(t)$ can be uniquely specified. A number of p.d.f for $f(t)$ have been proposed and examined in existing studies. This study examines the following three p.d.fs for the vehicle holding duration model, i.e., 1) Weibull, 2) Log-logistic, and 3) Log-normal. The p.d.fs are shown as follows:

1) Weibull: $f(t)=\gamma t^{r-1}\exp(-\gamma\beta X)\exp\left\{-t^{\gamma}\exp(-\gamma\beta X)\right\}$ (8a)

2) Log-logistic: $f(t)=\dfrac{\gamma t^{r-1}\exp(-\beta X)}{\left\{1+t^{\gamma}\exp(-\beta X)\right\}^2}$ (8b)

3) Log-normal: $f(t)=\dfrac{1}{\sqrt{2\pi}\sigma t}\exp\left\{-\dfrac{(\ln t-\beta X)^2}{2\sigma^2}\right\}$ (8c)

where, β , γ , δ and σ are unknown parameters that can be estimated, and X is the vector of covariates (independent variables).

In our model, some covariates may change over time. For example, household characteristics could change, irrelevant to the vehicle replacement. Pendyala et al. (1995) showed that the relationship between vehicle ownership and income is not stable. In such case, if one wants to incorporate the changes of the covariate into the model, the time-varying covariates should be introduced. Let the interval "0 to t" be divided into N, nonoverlapping intervals, $t_0 < t_1 < ... < t_N$, where $t_0 = 0$ and $t_N = t$. The covariates are assumed to be invariant within each interval, but they may vary from one interval to another. The survival function (eq. (6)) is rewritten as follows:

$$S(t) = \exp(-\int_0^t h(t \mid X(t))du) \tag{9}$$

where, $X(t)$ denotes a time-varying covariate at time t. The time-varying covariates are modelled as a step function, with different values through several intervals between $t = 0$ and $t = t_N$.

$$X(t) = \begin{cases} X_0 & t < t_1 \\ X_1 & t_1 \le t < t_2 \\ X_2 & t_2 \le t < t_3 \\ \cdot & \\ \cdot & \\ X_N & t_N \le t \end{cases} \tag{10}$$

The survival function with the time-varying covariate $X(t)$ is expressed as follows:

$$S(t \mid X(t)) = \prod_{n=1}^{N} \frac{S(t_n \mid X_{n-1})}{S(t_n \mid X_n)} \times S(t \mid X_N) \tag{11}$$

It is already known that estimating a model without these censored observations leads to self-selection biases (Marubini and Valsecchi 1995). Therefore, in this study we incorporate left and right-censored spells to avoid selection biases (Klein and Moeschberger 2003). The following log-likelihood function for the survival function model incorporates the censoring observation:

$$L = \prod_{i \in NC} f(t \mid X(t)) \cdot \prod_{i \in RC} S(t \mid X(t)) \cdot \prod_{i \in LC} \frac{f(t \mid X(t))}{S(v \mid X(t))} \cdot \prod_{i \in LRC} \frac{S(t \mid X(t))}{S(v \mid X(t))} \tag{12}$$

where, NC, RC, LC, and LRC are the numbers of non-censored, right-censored, left-censored observations, and left and right-censored observations, respectively.

2.3.2 Copula-Based Multivariate Survival Models

A bi-dimensional copula is a function $C_\theta : [0,1]^2 \to [0,1]$ with the following properties:

1) $C_\theta(0,u) = C_\theta(u,0) = 0$ and $C_\theta(1,u) = C_\theta(u,1) = u$, for all $u \in [0,1]$, and
2) $C_\theta(\cdot,\cdot)$ is bi-increasing; that is, for all $u' > u$ and $v' > v$:

   $C_\theta(u',v') - C_\theta(u',v) \ge C_\theta(u,v') - C_\theta(u,v)$.

Let T and D be two random variables with $F^T(t)$, $F^D(d)$ as marginal distribution functions, and let C_θ be a bi-dimensional copula, then the function $C_\theta(F^T(t), F^D(d))$ is a cumulative distribution function. Thus, copula functions are used to re-define joint distributions using the given margins. For any pair of scalar random variables (T,D) with the joint distribution function F, there exists a copula function C_θ such that,

$$F(t,d) = C_\theta[F^T(t), F^D(d)]. \tag{13}$$

The copula function C_θ is unique if the marginal distribution functions $F^T(t)$, $F^D(d)$ are continuous. Here, θ is a dependency parameter, which characterizes the dependency between $F^T(t)$ and $F^D(d)$.

One can also define copula densities in the same way as one defines probability densities. Let the distribution of (T,D) be continuous. The differentiated form of Sklar's theorem splits the joint density of T and D, $f(t,d)$, into the product of marginal densities $f^T(t)$ and $f^D(d)$, and the copula density, $c_\theta(u,v) \equiv \partial^2 C_\theta(u,v)/\partial u \partial v$, becomes,

$$f(t,d) = f^T(t) f^D(d) c_\theta[F^T(t), F^D(d)]. \tag{14}$$

Because $F^T(t)$, $F^D(d)$ have uniform distributions, $c_\theta(u,v)$ is the density of $(F^T(t), F^D(d))$ at (u,v) and it is also the conditional density of $F^D(d)$ at point v given $F^t(T) = u$.

To estimate unknown parameters, the following log-likelihood function is adopted.

$$\ln L(\alpha,\beta,\theta) = \ln f^T(t;\alpha) + \ln f^D(d;\beta) + \ln(c_\theta[F^T(t;\alpha), F^D(d;\beta);\theta]) \tag{15}$$

Copula function can be generated in different ways, including the method of inversion, geometric method, and algebraic method. In this paper, we will pick up the Normal, Gumbel, Clayton, and Frank copulas, which are useful for bivariate data, and present several desired properties. These copulas have been widely used because of their mathematical tractability. A detailed description about copula models is given by Nelsen (2006). We will apply these copulas, and select the best copula based on goodness-of-fit index.

1) Normal copula: $C(u,v) = \Phi_n(\Phi_1^{-1}(u), \Phi_1^{-1}(v);\theta) \quad (-1 \le \theta \le 1)$ (16a)

2) Gumbel copula: $C(u,v) = \exp(-((-\ln u)^\theta + (-\ln v)^\theta)^{1/\theta}) \quad (1 \le \theta)$ (16b)

3) Clayton copula: $C(u,v) = \left[\max(u^{-\theta} + v^{-\theta} - 1, 0)\right]^{-1/\theta} \quad (1 \le \theta < 0 \quad or \quad 0 < \theta)$ (16c)

4) Frank copula: $C(u,v) = -\frac{1}{\theta}\ln\left(1 + \frac{(\exp(-\theta u)-1)(\exp(-\theta v)-1)}{\exp(-\theta)-1}\right) \quad (\theta \ne 0)$ (16d)

where, θ represents dependence parameter, Φ indicates the standard normal distribution function, Φ^{-1} is the functional inverse of Φ, and Φ_2 is the bivariate standardized normal distribution function with correlation θ.

In this paper, copula-based models are used to capture and explore the dependencies between vehicle holding duration and annual travelling distance. For a pair (T,D) (T: vehicle holding duration; D: annual travelling distance) with a joint distribution function F, the joint survival function is given by $S(t,d) = P[T > t, D > d]$. The margins of S are the functions $S(t,-\infty)$ and $S(-\infty,d)$, which are the uni-variate survival functions $S^T(t)$ and $S^D(d)$, respectively.

Here, Let X and Y be continuous random variables with copula C_{TD}. Let α and β be both strictly decreasing on Ran T and Ran D, respectively. Then:

$$C_{\alpha(T)\beta(D)}(u,v) = u + v - 1 + C_{TD}(1-u, 1-v). \tag{17}$$

Using eq. (17) and the copula function (14), the joint survival function $S(t,d)$ is given as follows,

$$\begin{aligned} S(t,d) &= 1 - F^T(t) - F^D(d) + F(t,d) \\ &= S^T(t) + S^D(d) - 1 + C_\theta(F^T(t), F^D(d)) \\ &= S^T(t) + S^D(d) - 1 + C_\theta(1 - S^T(t), 1 - S^D(d)) \\ &= \hat{C}_\theta(S^T(t), S^D(d)) \end{aligned} \tag{18}$$

where, function $\hat{C}$ is a survival copula of T and D.

3. DATA

The data used in this study surveyed as the revealed preference. The data was collected in October 2006 from households living in the Chugoku area (the largest city is Hiroshima City) in Japan. All the households were asked to answer questions about the household and individual attributes, and the attributes of owned passenger vehicles in the past 10 years (i.e., from 1996 to 2006).

(1) Household attributes: number of household members, number of owned passenger vehicles, residential characteristics, etc.
(2) Individual attributes: age, gender, driving license, occupation, vehicle use behaviour, daily activity participation, etc.
(3) Events of household: moving house, marriage, birth of baby, purchase of vehicle, etc.
(4) Vehicle attributes: make, engine displacement, manufacture year, total travel distance of both current and previous vehicles, etc.

As a result, questionnaires from 500 households with vehicles were collected. In this study, vehicle holding duration model is built using a survey data of household vehicle ownership behaviour occurring from 1996 to 2006. The exact holding durations of previous vehicles purchased and got out from 1996 to 2006 were observed. When a vehicle is observed at the start and end of a holding spell, there is no problem. However, if the vehicle was bought before the survey, the sample cannot be used since we do not have the characteristics of household and main user before 1996. Furthermore, if the vehicle is under use during the survey period, the exact holding duration cannot be observed. As shown in Table 1, the sample used in this study includes 757 vehicles, among which 372 vehicles (49.1% of the sample) were replaced or disposed from 1996 to 2006. On the other hand, 101 vehicles (13.3% of the sample) were purchased before 1996 (left-censored data). The remaining 284 vehicles (37.5% of the sample) were purchased after 1996, but were still used by households at the time of survey (right-censored data). This study incorporates Non-censored data, left-censored data, and right-censored data to avoid selection biases.

Table 1 - Samples with censored data

DATA	Number of samples (percentage)
Non-censored	372 (49.1%)
Left-censored	101 (13.3%)
Right-censored	284 (37.5%)

Existing research has classified vehicle type based on various vehicle attributes, for example, vehicle size (e.g., Hayashi *et al.*, 2002), model type (e.g., Choo and Mokhatarian, 2004), fuel type (e.g., Golob *et al.*, 1996; Kho, 2003), and automaker (e.g., Kho, 2003). In Japan, such vehicle type is usually classified based on engine displacement. Such classification is well known to vehicle users. Since different tax systems are applied to vehicles with different engine displacements, vehicle users in Japan are very sensitive to engine displacement when purchasing vehicles. Therefore, this study defines the alternatives of passenger vehicles based on engine displacement, considering that this category is directly related to evaluation of fuel consumption, emissions and effects of vehicle-related taxes. Exploring the choice behaviour of passenger vehicle types is important for both marketers and public policy makers, especially considering that more and more people are showing concerns about environmental issues. For the purposes of estimating the choice models presented in this paper, the following three choice alternatives are adopted, considering the influence of sample size on the model estimation. Table 2 shows the share of the alternative in the sample. Middle-sized vehicle are the majority of vehicle type (54.7%), and the share of small-sized and large-sized vehicle are 27.3% and 18.0% respectively.

- Alternative 1 (Small-sized vehicle): passenger vehicle with engine displacement equal or smaller than 660cc
- Alternative 2 (Middle-sized vehicle): passenger vehicle with engine displacement larger than 660cc and equal or smaller than 2,000cc
- Alternative 3 (Large-sized vehicle): passenger vehicle with engine displacement larger than 2,000cc

Figure 1 shows the annual vehicle traveling distance, calculated from vehicle holding duration and total traveling distance. The average traveling distance is 10,015 km/year. It is observed that half of all vehicles are operated for less than 10,000 km/year. The number of households whose annual traveling distance is around 10,000-12,500 is relatively large.

Figure 2 shows the vehicle holding duration for those replaced or disposed from 1996 to 2006. The average vehicle holding duration was 4.48 years. As shown in Figure 1, the majorities of durations are 2 years and 3 years.

Table 2 - Shares of vehicle types

Alternative	Number of samples (percentage)
Small-sized vehicle	207 (27.3%)
Middle-sized vehicle	414 (54.7%)
Large-sized vehicle	136 (18.0%)

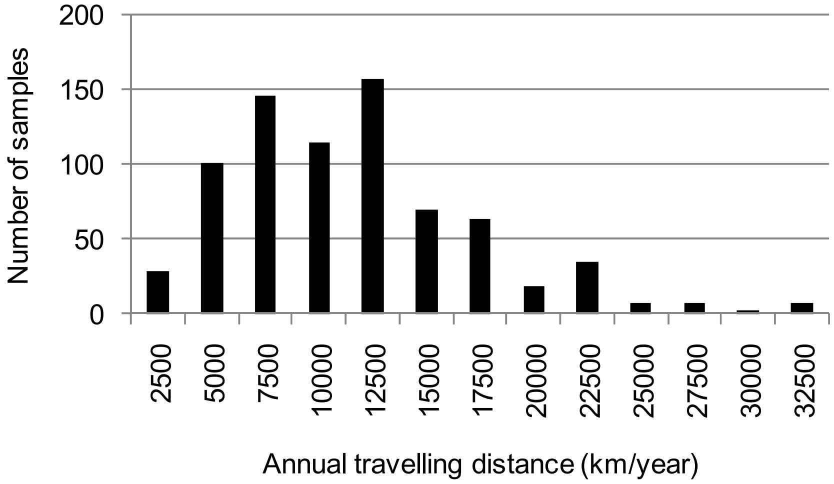

Figure 1 - Distribution of annual vehicle travelling distance.

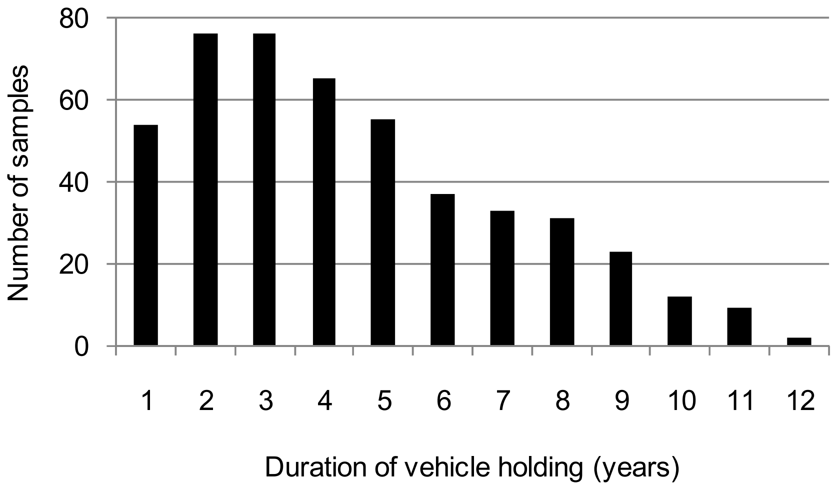

Figure 2 - Distribution of vehicle holding duration.

4. ESTIMATION RESULTS

4.1 Vehicle Type Choice

Estimation results of the vehicle type choice model are shown in Table 3. The adjusted McFadden's Rho-squared is 0.236, suggesting that the model is acceptable to represent household vehicle type choice in this study. Although fuel efficiency of the vehicle and taxes are used as one of the explanatory variables in existing studies, this study do not consider those variables. Because this study defined the alternatives by engine displacement, and these variables are strongly related with the engine displacement.

Concerning the influence of household attributes in estimated model, the coefficient of the average age of husband and wife has a positive sign and is statistically significant. This implies that older people prefer to own larger vehicles. For vehicle attributes, the composite variable "price / household income" and a number of passenger seats have negative values and are statistically significant, which indicate that inexpensive and smaller vehicles are preferred. The coefficient of annual travelling distance in the last year is positive sign and is statistically significant, which means the household with heavy user in travelling distance prefer to have middle-sized or large-sized vehicles.

Focusing on the dependency among the alternatives, the similarity coefficient between middle-sized vehicle and large-sized vehicle is statistically significant, but the others are not significant. These results indicate that there exists unobserved correlation between the middle-sized vehicle and large-sized vehicle, and applying the PCL model is validated. One of the plausible reasons of the correlation between the middle-sized vehicle and large-sized vehicle is the difference in the tax. The taxes of passenger vehicle include acquisition tax, auto tax and light vehicle tax, weight tax, and fuel tax. These tax systems are divided into 2 main classes; 1) tax system for light vehicle, and 2) tax system for middle-sized vehicle and large-sized vehicle. For example, acquisition tax is charged, when a vehicle is purchased. The tax rate is 5% for middle-sized vehicle and large-sized vehicle, but 3% for small-sized vehicle. Auto tax, which is a kind of property tax, is charged to the possession of a vehicle. Auto tax for a vehicle with less than 1,000cc is 29,500 yen/year, and for the vehicles with less than 3,000cc, an extra tax of 5,000 yen per 500cc has to be paid. In case that the displacement is larger than 3,000cc, the tax rises sharply, and the amount reaches 111,000 yen/year for a vehicle with more than 6,000cc. On the other hand, light vehicle tax is charged to the possession of a small-sized vehicle. The tax rate is 7,200 yen/year. One can see that there is a large gap between the tax systems for middle-sized vehicle and large-sized vehicle and the tax systems for small-sized vehicle.

4.2 A Joint Model of Annual Travelling Distance and Holding Duration

4.2.1 Selection of baseline hazard

Here, we compare several candidates of baseline hazard in order to find which underlying baseline hazard best fits annual travelling distance and holding duration based on the Bayesian Information Criterion (BIC) value. We examined three distributions: 1) Weibull, 2) Log-normal, and 3) Log-logistic distributions with respect to annual travelling distance and holding duration. The models are estimated for each dependent variable. Table 4 shows the baseline hazards for annual travelling distance and holding duration. It is clear that the Log-logistic distribution shows the best goodness-of-fit, followed by the Log-normal, and Weibull distributions. On the other hand, the hazard model of holding duration with Weibull distribution leads the best goodness-of-fit, followed by those with Log-normal, and Log-logistic distributions. Therefore, this study adopts the Log-logistic distribution for baseline hazard of annual travelling distance, and Weibull distribution for that of holding duration.

Table 3 - Estimation results of vehicle type choice model

Explanatory variable	parameter		t value
Household attributes			
Average age of husband and wife (M,L)	0.018	*	1.70
Number of workers (M,L)	-0.175		-0.57
Number of household members (M,L)	-0.116		-0.99
Number of household children (M,L)	0.216		1.62
Vehicle attributes			
Price/ Household income (S,M,L)	-2.377	**	-2.29
Number of passenger seats (S,M,L)	-0.361	*	-1.90
Vehicle use attributes			
Annual travelling distance in the last year (M,L)	0.687	**	4.50
Constant term			
Constant term (M)	-0.346		-0.65
Constant term (L)	-0.800		-1.42
Similarity coefficients			
ϕ_{12}	-9.162		-0.10
ϕ_{13}	-0.191		-0.47
ϕ_{23}	-2.646	*	-2.29
Number of samples	757		
Goodness-of-fit			
Initial log-likelihood	-831.650		
Converged log-likelihood	-631.973		
McFadden's Rho-squared	0.240		
Number of parameters	12		
Adjusted McFadden's Rho-squared	0.236		

Note
**: significant at the 10% level; **: significant at the 1% level*
1) Capitals in parentheses indicate the alternatives associated with this variable: (S)~660cc, (M) 661~2000cc, (L) 2001cc~.
2) Similarity coefficients defined as $\sigma_{jj'} = \exp(\phi_{jj'})/(1+\exp(\phi_{jj'}))$, where only $\phi_{jj'}$ is shown in this table.

Table 4 – Candidates of baseline hazards

	Converged log-likelihood	Number of coefficients	BIC
Baseline hazards of annual travelling distance			
Weibull	-548.014	13	591.105
Log-normal	-539.727	13	582.818
Log-logistic	-539.170	13	582.261
Baseline hazards of holding duration			
Weibull	-897.835	16	875.437
Log-normal	-918.031	16	904.501
Log-logistic	-929.054	16	912.963

$\mathrm{BIC} = -\ln(L_C) + 0.5p\ln(N)$
$\ln(L_C)$ is the log-likelihood vale at convergence
p is the number of coefficients
N is the number of samples.

4.2.2 Selection of copula function

Having selected the baseline hazards, here, we explore which type of copula is the most suitable to capture the interdependency between annual travelling distance and vehicle holding duration. Here, four copulas are considered as the candidates: 1) Normal, 2) Gumbel, 3) Clayton, and 4) Frank. Table 5 reports the BIC values for the CMS models. Comparison of the BIC values indicates that the CMS model with Clayton copula provides the best goodness-of-fit, followed by those with Normal, Frank, and Gumbel copulas. Moreover, fitness of CMS model with Clayton copula is higher than that of the conventional model assuming no interdependency between the annual travelling distance and vehicle holding duration. Thus, the proposed CMS model is outperformed over the conventional model. Hereafter, the CMS model with Clayton copula is applied to analyze the vehicle holding duration and annual travelling distance. And in the next section, the effects of covariates on annual travelling distance and holding duration are discussed.

Table 5 - Candidates of copula functions

	Converged log-likelihood	Number of coefficients	BIC
Normal	-1314.282	30	1413.722
Gumbel	-1361.425	30	1460.865
Clayton	-1313.370	30	1412.810
Frank	-1316.284	30	1415.724
Without copula	-1361.570	29	1457.698

4.2.3 Estimation Results with the Clayton Copula

The estimation results of the CMS model with Clayton copula, which has the highest goodness-of-fit, are shown in Table 6. The coefficient of copula is statistically significant and has a negative sign, which indicates the negative interdependence between annual travelling distance and holding duration. Therefore, the holding duration decreases as the annual travelling distance increases.

The estimated coefficients related to annual travelling distance, all of the coefficient estimates for household attributes, main-user attributes, and vehicle attributes have the expected signs. The coefficients of distances to the nearest rail station and supermarket have positive signs and are significant. These results seem reasonable since the households with low accessibility to station and supermarket tend to heavily use the vehicle. Moreover, the vehicles mainly used for commuting tend to have longer annual travelling distance since the coefficient of commuting distance has positive sign and is significant. The coefficient of wife dummy is negative, indicating that the wives drive shorter than the others. The signs of the coefficients of middle and large sized vehicle indicate that they are driving with longer distance than those of small-sized vehicles. The coefficient of running cost, which is one of the policy variables, has a negative sign and is significant. This result indicates that annual vehicle kilometers decrease as increases of fuel tax and gasoline price increase.

Observing the estimation results of the survival function related to holding duration, all of the coefficient estimates have expected signs. The coefficient in the number of household vehicles has a positive sign and is statistically significant, implying that the household with multiple vehicles holds each vehicle longer. The negative coefficient of income suggests that the households with high income tend to replace their vehicles much earlier than the others. Focusing on the coefficients of main-user attributes, the coefficient of commuting vehicle dummy indicates that vehicle used for commuting is held shorter than the other vehicles. The coefficient of main user age has a significant and positive. Older main users tend to hold their vehicles longer. The coefficient of logsum variable, which is calculated from the vehicle type choice model, has a significantly negative sign. The logsum variable is a measure of the price and quality of vehicle available on the market. The negative impact of logsum variable on duration indicates that holding duration decreases as the expected utility of the vehicle alternative increases. Therefore, the attractive vehicle alternative would shorter the holding duration.

Table 6 - Estimation results of CMS model

Covariates	Parameter		t-score
Interdependency between holding duration and annual travelling distance			
Type of Copula function	Clayton		
Copula parameter	-0.081	**	-39.5
Annual travelling distance			
Type of distribution	log-logistic		
Gamma	2.943	**	52.63
Constant	2.537	**	6.33
Household attributes			
Distance to the nearest rail station (m)	1.938	**	2.54
Distance to the nearest supermarket (m)	5.604	**	2.23
Household income	-0.019		-1.46
Main-user attributes			
Commuting distance to office (km)	2.133	**	2.40
Employed statement of main-user	0.010		0.12
Age of main-user	3.06.E-04		0.10
Housewife dummy	-0.380	**	-2.47
Vehicle attributes			
Middle-sized vehicle	0.516	**	5.15
Large-sized vehicle	0.785	**	5.37
Holding cost[a)]	-0.027	**	-3.32
Running cost	-0.013	**	-2.53
Holding duration			
Type of Distribution	Weibull		
Gamma	2.502	**	23.21
Constant	0.026	**	2.45
Household attributes			
Distance to the nearest rail station (m)	0.950		0.53
Distance to the nearest supermarket (m)	5.176		1.25
Number of workers	-0.002		-0.02
Number of license holders [a)]	0.090		0.56
Number of household vehicles [a)]	1.155	**	8.98
Household income	-0.160	**	-5.83
Main-user attributes			
Commuting vehicle	-0.612	**	-4.49
Age of main-user [a)]	0.028	**	3.20
Vehicle attributes			
Holding cost [a)]	-0.089	**	-4.47
Running cost [a)]	-0.047	**	-2.31
Middle-sized vehicle	0.357	*	1.86
Large-sized vehicle	0.524	*	1.84
Vehicle age	-0.258	**	-5.78
Expected utility			
Logsum from vehicle type choice [a)]	-0.986	**	-5.90
Number of samples	757		
Goodness-of-fit			
Initial log-likelihood [b)]	-1482.219		
Converged log-likelihood	-1313.365		

Note
**: significant at the 10% level; **: significant at the 1% level*
[a)]: Time-varying covariate
[b)]: log-likelihood with all coefficients to 0, except Gamma and constant.

5. SIMULATION RESULTS

Using the estimated models, we make a scenario simulation for possible policy measures. The simulations output the number of replacement, the changes in vehicle type choice share, expected annual travelling distance, and the amount of CO_2 emission under the different scenarios. First, "business as usual (BAU)" scenario is set to describe a situation without an policy (Case 0: BAU). The others scenarios with policies are set as follows:

Case 1: An increase in the running cost of 10% from 117 to 129 Japanese yen per litter.

Case 2: An increase in the holding cost. The amounts of the increasing holding costs HC_j for vehicle type j are calculated by eq. (19).

$$HC_j = (129 - 117) \times AVKM \ / \ FE_j \tag{19}$$

where, $AVKM$ indicates average annual travelling distance (i.e., equal to 10,015 km/year), and FE_j stands for the average fuel efficiency of vehicle type j (i.e., $FE_{small-sized}$ =12.7 km/l, $FE_{middle-sized}$ =10.1 km/l, $FE_{large-sized}$ =7.6 km/l)

Case 3: An increase in the purchasing cost. The cost is adjusted as identical to Case 1. The amounts of the increasing purchasing costs PC_j for vehicle type j are calculated by eq. (20).

$$PC_j = (129 - 117) \times AVKM \times AHD / FE_j \tag{20}$$

where, AHD represents average holding duration (i.e. equal to 4.48 years).

The variables related to household attributes, main-user attributes are assumed to remain unchanged, except the age of main-user. Moreover, we assumed that dependence structure between annual travelling distance and holding duration also unchanged (i.e. copula parameter is fixed).

For the simulations, first, we calculate the logsum value from the vehicle type choice model, which gives the attractiveness of vehicles available on the market. Then, the joint model of annual travelling distance and holding duration is estimated by using this logsum value, and the expected annual travelling distance and the expected survival probability of holding duration for each sample are calculated. For the households predicted to replace their vehicles (i.e., the expected survival probability in the holding duration is less than 0.5), the vehicle type choice model is used to calculate the probabilities for the replacing vehicles. These probabilities are summed up and combined with households which do not replace their vehicle, and the distribution of vehicle type share is calculated. Then, average fuel efficiency of whole samples (AFE) is estimated using the distribution of vehicle type share. Finally, the CO_2 emissions are calculated by multiplying the average fuel efficiency by the expected annual travelling distance ($EVKM$) as eq. (21);

$$\mathrm{CO_2\ emissions}(kg - CO_2 / year) = EVKM\,(km / year) / AFE(km / l) \times 2.3(kg - CO_2 / l)\ . \tag{21}$$

Simulation results are shown in Table 7. The results are summarized by the relative change to the BAU scenario (Case 0). These results therefore give predictions of the impacts of the respective policy measures only.

In case 1 in Table 7 shows that the increase in running cost leads to the decrease in the annual travelling distance, which fits our expectation. However, the rate of the decrease in the annual travelling distance is only 2.5%. On the other hand, forcing on the holding duration, it is considered that the influence of increasing of running costs on the holding duration appears in bi-ways from Table 6. Firstly, the running cost has a direct negative impact on the vehicle holding since the coefficient of running cost in hazard function related to the holding duration has a negative sign. Secondly, the running cost increases result in the decrease in travelling distance, which increases vehicle holding duration since they have negative interdependency. In our result, the direct effect dominates over the indirect effect as shown in table 7, holding durations also become shorter according to the increase of running cost, and more vehicles will be replaced by increasing on running cost. It is seemed because the data set in this study was collected in rural areas. Compared with large metropolitan areas, people's mobility in rural areas highly depends on private vehicles due to the lower level of services of public transportation systems and it is consequently difficult to reduce the annual travelling distance. Therefore, the rate of the decrease in the annual travelling distance is small, but number of vehicle replacements with the high fuel efficiency increase.

In case 2 about the increase on holding cost, the annual travelling distance is decreased, and the number of replacement is increased. The simulated results are similar with case 1. Moreover, from the comparison between case1 and case2, the change in share of vehicles is caused by the changes in the vehicle replacement rate. If the vehicle replacement rate is increased, the number of smaller vehicles is increased.

Focusing on the results of case 3, the annual travelling distance and vehicle share are not significantly changed.

From the perspective of environmental impacts, the running cost increase has the considerable impact on reduction of CO_2 emission, followed by holding cost. But, increase of purchase cost contributes to slight increase on CO_2 emission. This result indicates that the tax in vehicle use stage (i.e. increase on fuel tax) is the most effective in order to reduce the CO_2 emission.

Table 7 - Simulation results (relative change to the BAU scenario (Case 0))

	Case1 increase of running cost	Case2 increase of holding cost	Case3 increase of purchase cost
Expected annual traveling distance	- 2.50%	- 1.38%	- 0.04%
Expected holding duration	- 5.47%	- 11.03%	+1.46%
Number of vehicle replacements	+11.88%	+25.25%	- 1.98%
Share of small-sized vehicle	+2.22%	+3.11%	+0.00%
Share of middle-sized vehicle	+0.00%	+0.73%	- 0.24%
Share of large-sized vehicle	- 4.10%	- 8.20%	+0.82%
Total CO_2 emissions	- 2.80%	- 1.95%	+0.03%

6. CONCLUSIONS

There are various interdependencies related to the decisions on household vehicle ownership behaviours. Classifying these behaviours into the following three stages; choice of vehicle type, annual travelling distance, and holding duration, this study developed an integrated approach to accommodate them to the proposed model system. Annual travelling distance and holding duration are jointly modelled based on a Copula-based Multivariate Survival (CMS) model. Copulas are functions that calculate joint multivariate distribution functions by using their uni-variate marginal distribution function with limited number of dependency parameters. The uni-variate marginal functions are not necessarily the identical specification for all the dependent variables. A PCL model was further adopted to represent the vehicle type choice, and the expected utility of vehicle type choices from the PCL model is introduced into the CMS model to describe the dependence of annual travelling distance and holding duration on vehicle type choices. As the case study, a data of household vehicle ownership behaviour occurring between 1996 and 2006 was collected in the Chugoku region of Japan in 2006 and used to confirm the effectiveness of the developed integrated model. The data consists of four types of sub-samples in terms of the observation characteristics: i.e., non-censored, right-censored, left-censored, and left/right-censored samples. In this study, four popular copula functions, including Normal, Gumbel, Frank, and Clayton copulas, were empirically compared. As a result, it was found that the Clayton copula is the most suitable for the marginal function to describe both annual travelling distance and holding duration. The proposed CMS model with Clayton copula is superior to the conventional model without considering interdependency between the two behavioural aspects. The estimated coefficients showed that annual travelling distance and holding duration are significantly correlated with each other. It was also found that the holding duration decreases as annual travelling distance increases because of the negative interdependency between the two behaviours. Moreover, the coefficient of logsum variable in the vehicle holding duration was negative and significant. The negative impact on the vehicle holding duration indicates that it decreases as the expected utility in vehicle replacement increases.

In order to confirm the applicability of the proposed model as a policy evaluation tool, a simulation analysis was conducted to examine the effects of vehicle-related taxes on household vehicle ownership behaviour as well as CO_2 emissions. It was observed that the taxation in vehicle use stage is most influential to the reduction of CO_2 emissions. Moreover, an increase in tax in vehicle use stage can significantly cause the shift into smller engine vehicles. The increase in running cost on vehicle use was not simple. the impact of increase in running cost was not only to decrease in annual travelling distance, but also to increase the number of vehicle replacement into the smaller size. In predicting CO_2 emissions, the latter impact might be more important since the increase of the chare of smaller vehicles would contribute to effectively decrease CO_2 emissions.

The simulative analysis confirmed the effectiveness of the integrated model and its and applicability as a policy evaluation tool, while we also found remaining issues. Although the choice of vehicle types are linked to the travelling distance and the vehicle holding duration by using the logsum variable calculated from the utilities of vehicle types, the simultaneous modelling approach over these three behaviours is worth exploring. Furthermore, the copula-based modelling approach proposed in this study has a potential to accommodate such multi-dimensional behaviours with complicated dependence structures. The above modelling issues might be more useful when the model structure is expanded to be able to separately handle each vehicle, considering the interdependency among them, if the household holds two or more vehicles.

References

Bhat, C.R. and S. Sen (2006) Household vehicle type holdings and usage: An application of the Multiple Discrete-Continuous Extreme Value (MDCEV) model. Transportation Research Part 40B, 35-53.

Brownstone, D., Bunch, D. S., and Train, K. (2000) Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles, Transportation Research 34B, 315-338.

Choo, S., and Mokhtarian, .L. (2004) What type of vehicle do people drive? - The role of attitude and lifestyle in Influencing vehicle type choice, Transportation Research 38A, 201-222.

Chu, C. (1989) A paired combinatorial logit model for travel demand analysis, proceedings of the Fifth World Conference on Transportation Research 4, 295-309.

De Jong, G. (1996) A disaggregate model system of vehicle holding duration type choice and use, Transportation Research 30B, 263-276.

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

Golob, T. F., Bunch, D. S., and Brownstone, D. (1996) A vehicle use forecasting model based on revealed and stated vehicle type choice and utilization data, Journal of Transport Economics and Policy, 30, 69-92.

Hayashi, Y., Kato, H. and Teodoro R. V. R. (2001) A model system for the assessment of the effects of car and fuel green taxes on CO_2 emission, Transportation Research 6D, 123-139.

Hensher, D.A. (1985) An econometric model of vehicle use in the household sector, Transportation Research 19B, 303-313.

Klein, J.P. and Moeschberger, M.L. (2003) Survival Analysis (2nd ed.), Springer.

Koh, W.T.H. (2003) Control of vehicle ownership and market competition: Theory and Singapore's experience with the vehicle quota system, Transportation Research 37A, 749-770

Koppelman, F.S. and Wen, C.H. (2000) The paired combinatorial logit model: properties, estimation and application, Transportation Research 34B, 75-89.

Mannering, F., and Hamed, M. (1990), Occurrence, frequency and duration of commuter's work-to-home departure delay, Transportation Research 24B, 99-109.

Mannering, F., Kim, S., Barfield, W., and Linda, N. (1994), Statistical analysis of commuters route, mode and departure time flexibility, Paper presented at the 73rd Annual Meeting of Transportation Research Board, Washington D.C. (CD-ROM)

Marubini, E., and Valsecchi, M. G. (1995) Analysing survival data from clinical trials and observational studies, John Wiley & Sons.

Nelsen, R. B. (2006) An Introduction to Copulas (2nd ed.), Springer-Verlag, New York.

Pendyala, R. M., Kostyniuk, L.P., and Goulias, K.G. (1995) A repeated cross-sectional evaluation of car ownership, Transportation, 22, 165-184.

Yamamoto, T. and Kitamura, R.(2000) An Analysis of Household Vehicle Holding Durations Considering Intended Holding Durations, Transportation Research 34A, 339-351.