CHAPTER 4 METHODOLOGY

4.1 Sample

This study combined the motorcycles sampled in 2000 from the MOTC with the corresponding registration information for these sampled motorcycles from the VRS. Around 4 years of the observation period was conducted from the end of 1999 (the sampling time of MOTC) to February, 2004.

The MOTC's sampling survey was administered through a stratified and systematic sampling method, according to the registration area, motorcycle production year, and engine size recorded in the VRS on December 15, 1999 (MOTC, 2000). The effective sample size of the MOTC survey was 10,780 and sampling error was ±1% under a 0.95 confidence coefficient (MOTC, 2000). The survey mainly focused on motorcycle condition related to maintenance cost, frequency and purpose of use, running mileage, and basic information about the owners.

To study the association between duration variables and the determinants, three categories of covariates, including vehicle, usage, and aggregate attributes, were introduced to the duration models. In addition, the three types of variables were also applied into the logistic regression models to examine the odds of the initial ownership of a used/new motorcycle and the occurrence of a disposal/transfer event.

The vehicle and usage attributes were retrieved from both the MOTC's questionnaire and VRS records. In addition, based on the 23 districts for motorcycle registration in Taiwan, the aggregate variables indicating socioeconomic and motorcycle emissions inspection factors of a district were also introduced to further examine the influence of these aggregate predictors.

Vehicle attributes included used motorcycle, age of motorcycle at purchase, and motorcycle engine size. Motorcycle usage attributes included holder's sex, age, running mileage, maintenance costs, household motorcycle size, monthly income, and motorcycle registered in Greater Taipei or not. Of the aggregate attributes, two kinds of variables were used to represent socioeconomic index and motorcycle emissions inspection of a district respectively. The district aggregate variables can help establish the association between duration variables and these aggregate variables at a regional level. We chose the unemployment rate, motorcycle density, passenger car density, consumption propensity, and Engel's coefficient as the socioeconomic predictors. In addition, five emissions-related index for regional motorcycle inspection performance were included: the inspection rate, inspection station density, ineligibility rate, CO and HC emissions values of motorcycles averaged in a district.

4.2 Analytical Method

A great many time-dependent transportation issues using different types of duration models have been extensively reviewed (Hensher and Mannering, 1994). Hensher and Mannering indicated that several past transportation studies applied duration models, including parametric and semi-parametric (alternatively known as Cox proportional hazard) models, which focused on exploring the studies associated with time-to-accident, dynamic behavior of transportation demand, and holding duration of automobiles, etc. They also discussed some important issues about "heterogeneity", "state dependence", and "competing risks" relating to the formulation of a duration model.

To formulate our problems, the basic concepts, competing risk, semi-parametric and split-population duration models in survival analysis and logistic regression method are elaborated as follows.

4.2.1 Survival Analysis

4.2.1.1 Event, Censoring, and Survival Time

In longitudinal studies, the important feature is not only the event but also the time to that event. A survival analysis is a kind of statistical method aiming to study a group of subjects that experience specific events over a period of time. Specific events concerned in this study were the occurrence of motorcycle disposal and transfer records for motorcycle holding, and simply disposal records were concerned for measuring motorcycle age. The time that transpires until or between the occurrence of specific events is called the survival time (duration). Hence, a survival model is also called a duration model.

Three basic elements are needed in order to determine the survival time:

1. a time origin or starting point,
2. an ending event of interest, and
3. a measurement scale for the passing of time.

In addition, most survival analyses need to take into account the question of censored data. The censored data comprise only part of the complete information about survival time, but instead of eliminating these censored data the survival method can accommodate them into the analytical process. Three reasons in general for why censoring may occur:

1. a subject has not yet experienced the event before the study ends;
2. a subject is lost to follow-up during the study period;
3. a subject withdraws from the study for some other reasons.

In this study, the observation of the motorcycle age or holding duration respectively has two kinds of possible outcomes: event or censored states. All the censored data are right-censored because the initial time of the duration measured can be identified. The observation of a defined event for the duration variable has different meanings between motorcycle age and holding duration. In terms of motorcycle age, an event indicates only a sampled motorcycle undergoing the registration termination (i.e. disposal record) and a censored observation represent a motorcycle had not yet experienced the termination of registration within the observation period. However, in the analysis of motorcycle duration being owned, an event may indicate a sampled motorcycle underwent either a disposal or transfer record, and a censored observation had not yet experienced any disposal or transfer record.

To examine the survival experience in estimating motorcycle age or holding duration, the regression models including semi-parametric model (Cox regression) and parametric model (with Weibull hazard function) were established to explore the association between the duration variables and their determinants.

4.2.1.2 Definition of Hazard Function

Since motorcycle age or holding duration is a non-negative variable with non-normal distribution and with censoring nature, a duration (also called a hazard rate or hazard-based) model was deemed appropriate in describing its behavior. A variety of standard formulations for parametric and semi-parametric duration models are available in common statistical references (e.g. Kalbfleisch, 1980; Lawless, 1982; Allison, 1995; Le, 1997). Although the survival time may also be formulated as a dependent variable, the hazard rate is more conventional as it highlights the instantaneous event rate.

Given T, a continuous random variable with non-negative values, is defined as the duration variable for illustrating motorcycle age or holding duration (i.e. survival time), the cumulative distribution function *F(t)* means motorcycle age or holding duration less than *t* can be expressed as,

$$F(t) = \Pr(T < t) = \int_0^t f(x)dx, \forall t \geq 0 \quad (1)$$

where *f(x)* is the density function. Hence, the survivor function for motorcycle age or holding duration greater than or equal to *t* is written as,

$$S(t) = 1 - F(t) = \Pr(T \geq t) = \int_t^\infty f(x)dx, \forall t \geq 0 \quad (2)$$

The corresponding density function *f(t)* can be written as,

$$f(t) = \frac{dF(t)}{dt} = -\frac{dS(t)}{dt} = \lim_{\Delta t \to o^+} \frac{\Pr(t \leq T < t + \Delta t)}{\Delta t}, \forall t \geq 0 \quad (3)$$

In this study, the hazard rate is calculated for motorcycle disposal or transfer events. The hazard function *h(t)* is defined as an individual's instantaneous rate of transition from one state to another, given that the transition has not yet occurred before time *t*. The mathematical equation for the hazard function is expressed as,

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{\Pr(t \leq T < t + \Delta t)}{\Delta t \times \Pr(T \geq t)} = \frac{f(t)}{S(t)}, \forall t \geq 0 \quad (4)$$

To construct a regression model, it is necessary to define, in advance, the density, survivor, and hazard functions conditional on independent variables, denoted by $f(t|X,\beta)$, $S(t|X,\beta)$, and $h(t|X,\beta)$, respectively. The duration variable T denotes the length of time represented by motorcycle age or the time span of motorcycle being held, which was recorded in the VRS. The independent variable *X,* which is conditional on these functions, can be introduced to establish a regression relationship. To estimate the corresponding parameters of the independent variables, the general likelihood function form for the parametric models is given by:

$$L(\beta) = \prod_{i=1}^{n} \left[f(t_i | X, \beta) \right]^{C_i} \left[S(t_i | X, \beta) \right]^{(1-C_i)} \tag{5}$$

where n is number of observations and C_i is an indicator variable that equals 1 when the i^{th} observation of motorcycle holding termination occurs by time t_i and 0 otherwise. That is, the likelihood function is comprised of $f(t|X,\beta)$ for completed durations and $S(t|X,\beta)$ for censored observations. Therefore, the density function captures the fact that the observation experiences a motorcycle termination event by time t, whereas the survivor function catches the fact that the duration of the observation is at least as long as the censoring point, time t. In this study, duration was completed for the holders who terminated their motorcycle registration, and censored for those holders who did not, during the observation period.

Various regression forms for duration models may be assumed to describe the pattern of hazard rate as a function of time according to the assumed relationship between time and hazard. Two different kinds of logics are postulated in the respective parametric and semi-parametric models. For the parametric models part, a specific relationship between the hazard rate and time should be assumed, such as the exponential formulation with a constant hazard, Weibull with a monotonically increasing or decreasing, and lognormal with a first increasing and then decreasing form along with time. However, the semi-parametric model such as Cox regression model assumes the correct functional form can usually not be acquired and Cox model is robust to approximate the results of a correct parametric model without needing to assume a specific functional form in advance (Kleinbaum, 1995).

4.2.1.3 Competing Risk Approach

As previously mentioned, the termination of motorcycle holding were classed into two kinds of events: disposal and transfer. Motorcycles being disposed of may usually represent that only little residual value is left for these scrapped motorcycles, while for motorcycles being transferred, they still have higher values for sales. Hence, it is speculated that these two events have different natures considered by the owners who terminated their holding. That is, the two types of events reveal two competing risks for the owners who ended their motorcycle ownership.

Competing risk survival analysis using a modeling technique for latent variables assumes that only the shortest time among different competing risks can be observed. In this study, a disposal event observed means no transfer event will happen, and vice versa. Assume that T_j is the latent random variable for the survival time of risk j（j=1,2,…,r）among r different competing risks (r=2 in this study). Competing risk analysis introduces the individual hazard, survivor, and density function for each risk respectively (i.e. $h_j(t|X_j,\beta_j)$, $S_j(t|X_j,\beta_j)$, and $f_j(t|X_j,\beta_j)$).

The occurrence of different competing risks shows that there might exist either the association or independence among these risks. This study assumes a conditional independence between the two types of events due not only to this hypothesis with easy operation characteristic but also the lack of prior empirical studies regarding association. The conditional independence between risks define first that the overall hazard function is the summation of the individual hazard function for different r risks:

$$h(t|X,\beta)=\sum_{j=1}^{r} h_j(t|X_j,\beta_j) \tag{6}$$

Following the latent variable assumption that only the shortest time can be observed and provided that the jth risk at time t_i occurs for a motorcycle i, this means there will not be any risk j'（j'≠j）happened other than the jth risk before time t_i. Hence, the likelihood function contributed by the individual sample of the i^{th} motorcycle is expressed as:

$$L_i = f_j(t_i|X_{ij},\beta_j)\prod_{j'\neq j}^{r} S_{j'}(t_i|X_{ij'},\beta_{j'}) \tag{7}$$

If we further apply $h_j(t_i)S_j(t_i)$ to replace $f_j(t_i)$, the above equation reduces to:

$$L_i = h_j(t_i|X_{ij},\beta_j)S(t_i|X_i,\beta) \tag{8}$$

Therefore, the likelihood function caused by the jth risk is given by:

$$L(\beta_j)=\prod_{j=1}^{r}\prod_{i=1}^{n} h_j(t_i|X_{ij},\beta_j)^{C_{ij}} S_j(t_i|X_{ij},\beta_j) \tag{9}$$

where n is number of observations, C_{ij} is an indicator variable that equals 1 when the risk j occurs by time t_i and 0 when the other risks occur or the sample censored for observation i. Therefore, a competing risk approach with conditional independence assumption has kept the concerned risk as the event and reduced the other competing risks to censored observations while applying a survival analysis.

4.2.1.4 Cox Regression Model

The Cox regression model is comprised of two parts: the baseline hazard and an exponential component. Since the baseline hazard part is unspecified, Cox regression belongs to the semi-parametric class of models. Its functional form is expressed as

$$h(t|X,\beta) = h_0(t) \times e^{X \bullet \beta} \tag{10}$$

where $h_0(t)$ is the baseline hazard function and X and β are vectors of the independent variables and their corresponding parameters.

The estimated parameters $\hat{\beta}$ are obtained by maximizing the joint probability of the partial likelihood function $L(\beta)$,

$$L(\beta) = \prod_{q=1}^{k} \left(e^{X_{(q)} \bullet \beta} \Big/ \sum_{l \in R_q} e^{X_l \bullet \beta} \right) \tag{11}$$

where k is the number of events during the observation period; R_q is the set of motorcycles at risk (the risk set) at the q^{th} rank of time ($\leqq k$) for the events, which represents the set of motorcycles that have not experienced any events before the q^{th} rank of time; $X_{(q)}$ is the vector of independent variables corresponding to the event occurring at the q^{th} rank of time (number q in parenthesis represents the rank order of time); and X_l is the vector of independent variables corresponding to the l^{th} set of motorcycles in R_q.

The estimated parameters $\hat{\beta}$ derived from the above function, while not fully statistically efficient, are consistent and in large samples also asymptotically normal and unbiased.

When comparing the relative hazard contribution of two distinct independent variable vectors, it is conventional to use the hazard ratio. This is expressed by equation (12), which is constant since $h_0(t)$ cancels out of the equation (the proportional hazard assumption).

$$HR = \hat{h}(t|X_1, \hat{\beta}) \Big/ \hat{h}(t|X_2, \hat{\beta}) = e^{(X_1 - X_2) \bullet \hat{\beta}} \tag{12}$$

where X_1 and X_2 stand for the two values of the independent variable vector considered. If HR is significantly greater than 1, then the conditions described by X_1 are more likely to terminate a motorcycle's holding than the conditions given by X_2. In other words, the motorcycle life expectancy is shorter under X_1.

4.2.1.5 Split-population Duration Model with Weibull Hazard Function

Some of the censored holdings, identified by the registration records, may actually have been terminated, but not registered in the VRS. At least 11.8% of registered motorcycles were no longer in the possession of their registered owners; this could not be adequately demonstrated by the VRS, however (MOTC, 2000).

As a result, applying the registration history of a motorcycle from the VRS may have overestimated average holding duration or scrappage age, since some of the motorcycle owners had ceased ownership, while the corresponding registration records had not been changed. The specific type of heterogeneity for the holding duration, which partly came from inaccurate registration records in the VRS, revealed that a special statistical method for formulating this problem should be considered. Thus, to revise the specific heterogeneity of these censored data, a split-population duration model was employed to reduce any possible bias.

An implicit assumption made in standard duration models is that the event will eventually occur, which in this context implies that all motorcycle holders will eventually terminate their holding in the long run. However, the duration concerned was obtained from the VRS records, where some holders may not have cancelled their motorcycle registration when holding had actually been terminated. Estimation errors, produced by these partially incorrect VRS records, were corrected using a split-population duration model that took into account the possibility that the event may never occur (Schmidt and Witte, 1989). This correction appears appropriate, since a portion of the holders who ceased holding motorcycles, may never have terminated their registration status in the VRS. Thus, the probability that a motorcycle holder will eventually register termination of ownership in the VRS was postulated to be less than one. The split-population model estimated the instantaneous rate of the motorcycle population that would eventually terminate, as well as the subpopulation that would never experience a termination event.

The standard duration model assumes that censored observations will eventually terminate their motorcycle registration, whereas the split-population duration model allows the possibility that some censored observations may never encounter a registration termination. To illustrate the split-population duration model, let U be an unobservable variable that equals 1 for those who will eventually terminate their motorcycle registration and 0 otherwise. Then,

$$P(U=1)=\delta, \qquad P(U=0)=1-\delta \tag{13}$$

where δ is the “split-population parameter” that denotes the probability of eventual termination of motorcycle registration. If $\delta<1$, then a proportion of the censored observations will never experience this termination. Therefore, a split-population may be viewed as two different parts of a survival experience. For the first part, if a holder has terminated his/her motorcycle registration, we have $T=t$ and $C_i=1$, the appropriate density for such a holder is therefore,

$$P(U=1)f(t|X,U=1)=\delta f(t|X,U=1) \tag{14}$$

For the second part, a censored observation $C_{\mathrm{i}}=0$ reveals simply that ownership of the motorcycle has not ended during the observation period; we cannot, however, identify whether this censored data will never experience that event, or whether the observation time is simply not long enough to capture the event. Thus, we may consider the following two possibilities, (a) that this motorcycle will never undergo a change in ownership registration or (b) that registration of the motorcycle would have been terminated if it had been followed for a longer time. Specifically, this censored condition can be expressed as,

$$P(U=0)+P(U=1)P(T>t|X,U=1)=1-\delta+\delta S(t|X,U=1) \tag{15}$$

Therefore, the likelihood function for the split-population duration model consists of expressions (14) and (15) above for completed and censored durations respectively. The specific likelihood function form can be determined once the hazard function is parameterized. In this study, we speculated that the chance of a motorcycle being disposed of or transferred will increase over time. As a result, a Weibull hazard function is appropriate to our question, as it possesses a flexible property that can determine the estimated model with a constant, decreasing, or increasing hazard rate.

The Weibull hazard function and survival function is respectively given by:

$$h(t)=\lambda p(\lambda t)^{p-1} \tag{16}$$

$$S(t)=\exp(-(\lambda t)^{p}) \tag{17}$$

where $\lambda=\exp(X\cdot\beta)$, i.e. β is the estimated parameter of the independent variable vector *X*, and *P* is the "shape parameter" that determines how the hazard rate is dependent with time. If *P*=1, the Weibull model reduces to the exponential model with a constant hazard rate. If *P*>1, the model demonstrates an increasing hazard rate over time; otherwise *P*<1 demonstrates a decreasing hazard rate. We assumed in this study that the hazard rate of a motorcycle holding being terminated would reveal a growing rate over time (i.e. *P*>1). Limited by a one-wave investigation from the MOTC's sampling inventory, fixed covariates were introduced into the duration models in this study.

After specifying the Weibull hazard form, the likelihood function for a split-population model is expressed as,

$$L(\beta) = \prod_{i=1}^{n} \left[\delta f(t|X,\beta)\right]^{C_i} \left[1-\delta+\delta S(t|X,\beta)\right]^{(1-C_i)}$$

$$= \prod_{i=1}^{n} \left[\delta\lambda p(\lambda t_i)^{p-1} \exp(-(\lambda t_i)^p)\right]^{C_i} \left[1-\delta+\delta \exp(-(\lambda t_i)^p)\right]^{(1-C_i)} \qquad (18)$$

The parameter estimates, $\hat{\beta}$ and $\hat{\delta}$ may be obtained by maximizing the above likelihood function. The split parameter $\hat{\delta}$ allows the probability of eventual termination of the motorcycle holding to be less than one. If the estimated $\hat{\delta}$ is significantly less than 1, the split-population model could validate our speculation that a portion of the VRS records did not report the actual holding time so as to inflate the average holding duration or scrappage age; otherwise it converges to a standard hazard model.

The LIMDEP Version 7.0 software (Greene, 1998) was applied to estimate the split-population duration model. However, the estimation of parametric duration models in LIMDEP uses the formulation of the log-survival time rather than the log-hazard form. To compare with the hazard form of Cox regression model consistently, the estimated parameter and hazard ratio for each variable has been transformed into the log-hazard form. The log-hazard model is expressed as:

$$\log h(t) = (P-1)\log t + X \cdot \beta \qquad (19)$$

The log-survival model, the exact equivalence of the above log-hazard form, is:

$$\log T = X \cdot \beta^{*} + \sigma\varepsilon \qquad (20)$$

where T is the random survival time that conform with a Weibull distribution, ε is the random disturbance term that conform with a extreme value distribution, and σ is the scale parameter. The equivalence of the estimated parameters between the two models can be expressed as $\beta = -\beta^* / \sigma$ and $P = 1/\sigma$ (Allison, 1995).

4.2.2 Logistic Regression Method

The logistic regression model was employed in this study to estimate the relative likelihood between disposal and transfer competing risks for terminating the motorcycle ownership and between used and new motorcycle purchased associated with specific determinants. This regression model is widely used if the dependent variable is dichotomous. It is one form of statistical model called "generalized linear model" with a logit (also called "log odds", i.e. ln *P/*(*1-P*)) link function. Logistic regression allows us to predict a binary outcome (i.e. disposal and transfer termination) from a set of explanatory variables that may be continuous, categorical, or a mixture of the two.

The dependent variable in logistic regression is dichotomous; that is, it is assumed to follow a Bernoulli process. Therefore, it takes the value 1 with a probability *P* of an event occurred, and the value 0 with probability 1-*P* of an event not occurred. The form of the logistic regression equation is (McCullagh & Nelder, 1989):

$$\ln\left[P(X)/(1-P(X))\right] = \alpha + X \cdot \beta \tag{21}$$

where *P(X)* is the probability of terminating a motorcycle with a disposal event or holding a second-hand motorcycle at the initial buying in the respective two models, which is a function of the vector of determinant factors *X*; α is the constant of the equation, and β is the vector of the factor coefficients.

The coefficients of the logistic model can be obtained by using the maximum likelihood estimation method. A Wald test is usually used to test the statistical significance of each coefficient in the model. The squared Z statistic yields a Wald statistic of asymptotic chi-square distribution with one degree of freedom, that is:

$$Wald = \left(\hat{\beta}_i / SE(\hat{\beta}_i)\right)^2 \sim \chi_1^2 \tag{22}$$

where $\hat{\beta}_i$ is the ith estimated coefficient and $SE(\hat{\beta}_i)$ is the standard error of the ith estimated coefficient.

To interpret the model conveniently, logit (i.e. ln *P*/(1-*P*)) can be converted easily into a statement about odds ratio (O.R.) of the dependent variable simply by using the exponential function. For example, if the x_i variable increases one unit while holding the remainder variables constant, the O.R. of these two levels for x_i will be exp($\hat{\beta}_i$) and the 95% confidence interval (C.I.) for O.R. will be exp($\hat{\beta}_i \pm Z_{0.95}$*SE($\hat{\beta}_i$)).

4.3 Model Specification and Hypotheses

Three types of covariates, including vehicle, motorcycle usage, and aggregate attributes, were introduced to the two kinds of duration models (i.e. motorcycle scrappage age and holding duration) in this study. The same variables for vehicle and motorcycle usage attributes were used in the respective motorcycle age and holding duration models. The aggregate attributes examined the association with duration variables at a regional perspective. The aggregate variables concerned in the motorcycle disposal age and holding duration models, however, had a little difference. The difference came from the hypothesis that regional motorcycle life span, although not holding duration, might be associated with the district's emissions control policies. Therefore, predictors indicating district's performance of emissions inspections were additionally added in examining motorcycle disposal age.

Expectations of the relationships between the independent variables and the duration (or inversely, the hazard rate) came partly from the aforementioned research on automobiles (Gilbert, 1992; De Jong, 1996; Yamamoto and Kitamura, 2000). In addition, we also raised our own speculations for specific variables for the purpose of exploration. Since two kinds of duration models for motorcycle disposal age and holding duration may have different association hypothesis, the variable descriptions and their associated assumptions are individually illustrated according to the two kinds models as follows.

1. Vehicle attributes

(1) Used motorcycle

This variable was based on whether a holder had obtained a second-hand motorcycle at the initial holding. This was identified by whether the sampled motorcycle revealed any transfer records in the VRS before December 15, 1999. Provided at least one transfer record was present, the variable represented a holder having acquired a used motorcycle; otherwise the motorcycle was deemed to have been new at the time of purchase.

A used vehicle at purchase revealed a higher probability of being replaced sooner (De Jong, 1996; Yamamoto and Kitamura, 2000), because of the lower reliability and higher repair cost (De Jong, 1996). Similarly, since a used motorcycle was assumed to have a relatively higher possibility of attrition and maintenance costs compared to a new one, the holding duration of a used motorcycle was expected to be shorter (i.e. the hazard rate of terminating ownership would be higher).

However, in terms of the whole motorcycle scrappage age, this association might not be obvious. Unlike positively associated with the hazard of the holding duration, the scrappage hazard contributed by a used motorcycle was indeterminable.

(2) Age of motorcycle purchased

In addition to the used status of a motorcycle, the age of a used motorcycle at the initial holding may also influence the holding duration or motorcycle scrappage age. This variable was measured by the time between the original issue time of motorcycle plate and the time of the "latest transfer", if any, before the sampling time of the MOTC survey. If no transfer record existed, the motorcycle's age purchased was set to zero. This variable was used as one of the independent variables in the duration models for ownership duration and disposal age respectively. Also, it was put as a dependent variable to study the possible determinants that affected the age of a used motorcycle at the initial purchase.

Car age at purchase was found to be positively associated with the hazard of being replaced (De Jong, 1996; Yamamoto and Kitamura, 2000). Therefore, we assumed that the older a used motorcycle at purchase, the higher the instantaneous termination of that motorcycle ownership. However, this association with the entire motorcycle scrappage age by disposal was not so obvious and thus assumed to be indeterminable.

(3) Engine Size

As previously mentioned, both the import and issue of certificate for motorcycles with engine capacity over 150cc has been prohibited since 1980 due to the oil crisis. This ban was not lifted until July, 2002 because of Taiwan joined the WTO. Consequently, the majority of motorcycles in Taiwan are mopeds and light motorcycles less than 150 cc. In general, motorcycles over 150cc have a longer holding duration or life span not only because of their higher value, but also because of their scarcity resulting from the long-standing registration prohibition.

Since a higher engine size may represent a higher motorcycle replacement cost, we speculated that the higher the engine size, the lower the hazard for terminating holding or disposing of a motorcycle. The variable for engine size was classified into three categories: 50cc or under (i.e. mopeds); 51-150cc (light motorcycles); and those over 150cc.

2. Motorcycle usage attributes

(1) Holder's sex

Different sex may have diverse lifestyle, taste, and sense of value, thus influencing preference towards the time when durable goods may be replaced or disposed of. According to De Jong (1996), female owners had a higher probability of replacing their vehicle sooner than males. Hence, it was assumed that female motorcycle owners had a higher hazard rate of ending ownership in terms of holding duration, but no prior assumption was made about motorcycle scrappage age in this study.

(2) Holder's age

Different age cohort may also have different preference towards the time when durable goods may be replaced. Vehicles used by older primary users revealed a longer ownership period (Yamamoto and Kitamura, 2000). Older people tended to possess cars for a longer period, possibly due to the formation of stronger habits in the duration model with time-varying covariates indicated by De Jong (1996), but the inverse results was shown by his conventional duration models. In this study, younger people were hypothesized to be more likely to replace their motorcycles, and thus to have a raised hazard of ending ownership both for holding duration and motorcycle scrappage age. For comparison, the age of these sampled holders were adjusted to the time the event had occurred or was censored.

(3) Running mileage

Running mileage of a motorcycle was considered to be positively connected with its wear and tear condition. Higher vehicle miles traveled showed a higher hazard for terminating vehicle ownership (De Jong, 1996; Yamamoto and Kitamura, 2000). Similarly, vehicles with higher odometer readings also revealed a higher terminating hazard (Gilbert, 1992). The running mileage per week of the MOTC's sampled motorcycles was employed to examine the association to both holding duration and motorcycle age. As expected, a higher running mileage was likely to increase attrition and risk of motorcycle damage so as to decrease the holding duration or motorcycle scrappage age.

(4) Maintenance costs

By applying the annual self-reported maintenance costs of the sampled motorcycles, this study was able to indirectly observe the wear and tear of the motorcycles. This variable was measured by taking the logarithm of the yearly maintenance costs of the sampled motorcycles. We speculated that a motorcycle with a higher maintenance cost was more likely to arouse its holder's intention to replace or dispose of it, which increased the hazard of ending ownership.

(5) Motorcycle size of household

On average, 1.78 motorcycles were owned per household at the end of 2003. Possessing additional motorcycles for a sampled motorcycle holder in his/her household may influence the sampled motorcycle holding duration or life span. The household fleet size was negatively correlated with the holding duration of the observed vehicle (De Jong, 1996; Yamamoto and Kitamura, 2000). That is, as the number of vehicles in a household increased, the shorter the holding duration (or the higher the hazard rate) of the observed vehicle. This may be produced by the income effect that more cars in the household indicating a higher consumption propensity led to a sooner replacement on a vehicle holding.

This variable was classified into three classes in this study: the only motorcycle, two motorcycles, and three motorcycles or more in the holder's household. Analogical speculation from previous automobile studies, the less the motorcycle being owned in a household, the lower the hazard rate of terminating holding of the sampled motorcycle.

(6) Car size of household

Jou et al. (2004) found the existence of the substitute relationship between car and motorcycle both for their ownership and usage in a household. We speculated more cars for a sampled motorcycle holder in his/her household may extend the sampled motorcycle holding duration or life span and thus reduce the hazard rate of motorcycle terminating holding. Three classes of household's passenger car size were classified: none, one car, and two cars or more in the holder's household.

(7) Monthly income

Yamamoto and Kitamura (2000) indicated that high personal income reduced the vehicle holding duration. Gilbert (1992) also found that high income decreased the vehicle ownership duration by replacing a new car or by disposing of it, but increased the ownership duration by replacing a used car. Therefore, lower personal monthly income by the motorcycle holders was assumed to extend their motorcycle holding duration or scrappage age due to their relatively conservative consumption behavior. Monthly income was divided into binary levels: less than 30,000 NT dollars and 30,000 NT dollars or more.

(8) Greater Taipei

Having the only mass rapid transit system in Taiwan, Greater Taipei covering Taipei city and county has been concerned about the difference of motorcycle holding behavior with other districts. No previous research being conducted on this issue, and thus no assumption was assumed in advance about whether motorcycles registered in Greater Taipei associated with the ownership duration or motorcycle disposal age.

3. Aggregate attributes

Different administrative divisions may possess diverse characteristics regarding the standard of living, the traffic composition, and transportation management policies, etc. The aggregate factors in these districts induced our concerns over the association between these factors and motorcycle holding duration or disposal age at a regional level. Motorcycles sampled from the MOTC's inventory, covering all of Taiwan's twenty-three administrative areas, offered us the opportunity to explore how the socioeconomic index and motorcycle emissions inspection performance of a district may influence the duration variables.

Various types of socioeconomic variables have been introduced into different studies according to each author's discretion (Gilbert, 1992; De Jong, 1996). In addition to the assumption drawn from past research, that a high unemployment rate caused a decline in the hazard for vehicle disposal (Gilbert, 1992), we introduced another four socioeconomic variables for exploration purposes. These included the motorcycle density, passenger car density, Engel's coefficient, and the consumption propensity measured in these twenty-three administrative areas. Since these five aggregate covariates did not reveal substantial variations across the observation period, our concern over the influence of these covariates on duration variables mainly came from the variations across areas, but not across time. The average values of these socioeconomic variables during 1999-2003 were employed to relate to the period that the events were observed (Directorate-General of Budget, Accounting and Statistics, Executive Yuan, 2006).

(1) Unemployment rate

The average unemployment rate in the respective areas according to motorcycle registration represented the general economic condition of these areas. It was assumed that a soaring jobless rate would, on average, make consumers more conservative in their consumption behavior, then decreasing the holder's intention to replace or dispose of their motorcycles in these areas. Thus, we postulated that a higher unemployment rate in an area would lead to a declining average hazard ratio for motorcycle holding termination.

(2) Motorcycle density

This variable was measured via the ratio of numbers of motorcycles owned per thousand persons within the motorcycle registration area. An area having a high motorcycle density revealed that motorcycles were more likely to be the principal transportation mode within that area, with, on average, a higher likelihood of intensive usage, which would result in a higher possibility of ending motorcycle ownership.

(3) Passenger car density

Similar to the above variable, this variable was defined as the ratio of numbers of passenger cars owned per thousand persons within each of the twenty-three areas. Since passenger cars may have a partial substitution effect on motorcycles, areas with a higher ownership rate of passenger cars were more likely to reduce the average usage frequency of motorcycles. Therefore, this variable was expected to have a negative influence on hazard rate.

(4) Consumption propensity

Consumption propensity in an area was an index measuring the average intensity of purchasing goods in terms of household level. It was defined as the proportion of a household's consumption expenditures over the average disposable income of that household. We speculated that the stronger the consumption propensity, the more likely the decision to replace or dispose of a motorcycle. That is, the expected influence of this variable was positively connected to the hazard rate for motorcycle ownership termination.

(5) Engel's coefficient

The Engel's coefficient of an area was defined as the percentage of average total food-and-drink expenditures over the average total consumption expenditures of households in an area. A higher coefficient meant that on average the households in this area spent more on food-and-drink, thus spending less on durable goods such as buying motorcycles. We speculated that a higher Engel's coefficient in an area discouraged the purchase of durable goods, thus extending the average holding duration or scrappage age of motorcycles in that area.

(6) Motorcycle inspection performance

Although the emissions standards for motorcycles are a national standard, the actual emissions inspections are implemented by the Environmental Protection Bureau within each local administrative district. Each local district shows diverse averages for the CO and HC emissions levels of inspected motorcycles, as well as diverse inspection rates and ineligibility rates. In addition, the density of the stations for motorcycle inspections differs substantially among the different districts. This raises the concerns as to whether these differences in regional inspection performance for motorcycles affect the disposal age of motorcycles at a regional level.

It was assumed that motorcycle inspection performance might be associated with motorcycle scrappage age of a district. The inspection performance was treated two different ways separately. First, the average inspection rate, inspection station density, and ineligibility rate were applied as continuous district variables (the average levels of CO and HC emission were excluded from the duration models because they are highly correlated with the ineligibility rate). Second, we created a dummy variable to indicate membership in one of the two groups defined by the cluster analysis.

The average inspection rate was measured by the participation percentage of the annual mandatory inspection for motorcycles emissions in each administrative district. The inspection station density was calculated by the number of approved inspection stations per 10 thousand registered motorcycles in each district. In addition, the average ineligibility rate was measured by the percentage of motorcycles failing to pass the applicable standards for either CO or HC emissions, while the average CO and HC emissions were measured by the average CO values (%) and HC values (ppm) for motorcycles inspected by idle testing in each district.

The above mentioned inspection rate, inspection station density, and ineligibility rate were averaged as continuous variables among the 23 administrative districts during 1999-2003. It was assumed that the higher inspection rate and the lower ineligibility rate represented a higher inspection performance in a district, thereby increasing the disposal hazard of motorcycles in the district. The inspection station density, however, had no prior speculation. In addition, a cluster of districts showing good inspection performance was assumed to have higher disposal hazard as well.

The notation and description of these independent variables and the expected influence on hazard ratio by changing them on the two duration models are summarized in Table 3. In addition to the two main duration models, we also examined an additional duration model for the age of second-hand motorcycle purchased and two logistic regression models for dichotomous competing events and used/new motorcycle purchased respectively. As exhibited in Table 3, we have put much emphasis on the two duration variables and omitted the association expectations of other models.

Table 3 Expected Association of Independent Variables on Hazard Ratios of Two Duration Models

Variable notation	Variable description	Expected association	
		Scrappage age	Ownership duration
Vehicle attributes:			
Used motorcycle	1 if the sampled holder possesses a second-hand motorcycle, otherwise 0	**?**	+
Age of motorcycle purchased	Age of the sampled motorcycle age at the initial holding	**?**	+
CC(1)	1 if the motorcycle's engine size is less than or equal to 50cc, otherwise 0	+	+
CC(2)	1 if the motorcycle's engine size is in the range 51cc - 150cc, otherwise 0	+	+
Motorcycle usage attributes:			
Sex	1 if the holder of the sampled motorcycle is male, 0 if female	**?**	-
Holder's age	Age of the sampled motorcycle's holder	-	-
Running mileage	Running mileage per week of the sampled motorcycle	+	+
Maintenance costs	The logarithm of the yearly maintenance cost of the sampled motorcycle	+	+
Household motorcycle size(1)	1 if the household of the sampled motorcycle possesses the only motorcycle, otherwise 0	-	-
Household motorcycle size(2)	1 if the household of the sampled motorcycle possesses two motorcycles, otherwise 0	-	-
Household car size(1)	1 if the household of the sampled motorcycle does not possess any cars, otherwise 0	+	+
Household car size(2)	1 if the household of the sampled motorcycle possesses on car, otherwise 0	+	+
Monthly income	1 if the monthly income of the sampled holder less than 30,000 NT dollars, otherwise 0	-	-
Greater Taipei	1 if the sampled motorcycle registered in Taipei city or county, otherwise 0	**?**	**?**
Aggregate attributes:	(By district, averaged over the period 1999-2003)		
Unemployment rate	The average of annual unemployment rate	-	-
Motorcycle density	The average numbers of motorcycles per thousand persons	+	+
Passenger car density	The average numbers of passenger cars per thousand persons	-	-
Consumption propensity	The average consumption expenditure per household expressed as a percentage of the average disposable income	+	+
Engel's coefficient	The average food and drink expenditure per household expressed as a percentage of the average consumption expenditure	-	-
	(Continuous variables)		
Inspection rate	The average of annual inspection rate	+	**N.A.**
Ineligibility rate	The average of annual ineligibility rate	-	**N.A.**
Inspection station density	The average numbers of inspection stations per 10 thousands motorcycles	**?**	**N.A.**
	(Dummy variable)		
Inspection performance cluster	1 if the motorcycle is registered in a district belonging to the cluster with higher overall inspection performance, otherwise 0	+	**N.A.**

Note: 1. "?" represent expected association cannot be determined and "N.A." means "not applicable".
2. "Vehicle", "motorcycle usage", and "aggregate" attributes were retrieved from the VRS records, 2000's MOTC sampling survey, and officially published index, respectively.