An Overview of Parametric Hazard Functions: Applications in Vehicle Transaction Decision

Taha H. Rashidi (Ph.D. Candidate, Corresponding Author, thosse2@uic.edu)[1]
Abolfazl (Kouros) Mohammadian (Associate Professor, kouros@uic.edu) [1]

A Short Paper Prepared for the 3rd conference on Innovations in Travel Modeling

Abstract

Hazard models are moderately utilized in the transportation field while their applications in other areas such as medicine, political science and economics go back to more than fifty years ago. Hazard models are capable of modeling the failure time of a decision or an event, such as vehicle transaction and job type/location change decisions. Many specifications of the hazard models are not well-known while the primary application of hazard models is limited to a basic proportional continuous hazard formulation with continuous failure time and Weibull baseline hazard function. Recent advances including the application of accelerated hazard formulation is a major step forward, however such applications are limited and the proportional hazard formulation remains as the proffered hazard-based modeling approach. One burden holding back researchers from using more sophisticated hazard models could be attributed to the complexity of mathematical formulation of these approaches compared to the basic proportional hazard formulation. This paper aims to present some of the important specifications of hazard-based models that are less frequently employed in practical models while their application can advance the modeling results. In this study, the log-logistic baseline hazard which is a non-monotonic function is substituted with the monotonic Weibull function and the modified formulation is presented. The log-logistic and Weibull formulations are also presented not only in the continuous form but also the discrete hazard formulations for applications where the failures are observed in discrete time intervals. Finally, all four combinations of discrete and continuous formulations with Weibull and log-logistic baselines are discussed with the option of unobserved heterogeneity of type of gamma distribution. A case study of vehicle transaction type and timing decision at household level is finally discussed using the eight formulations presented in this paper.

[1] Department of Civil and Materials Engineering
University of Illinois at Chicago
842 W. Taylor St.
Chicago, IL 60607
Phone: 312-996-0962
Fax: 312-996-2426

Introduction and Background

A hazard model with continuous failure time formulation along with a Weibull baseline hazard and no embedded heterogeneity is a typical format of hazard models utilized in transportation and many other fields (Hensher and Mannering 1994, Yamamoto *et al.* 1999, Yamamoto *et al.* 2004, and Mohammadian and Rashidi 2007). However, it has been suggested that in a parametric hazard model, effectiveness of a non-monotonic baseline hazard (Yamamoto *et al.* 1997), a discrete formulation (Han and Hausman 1990) and an unobserved heterogeneity (Bhat 1996) should be verified and tested. Bhat (1996) studied a discrete hazard formulation doubled with Weibull baseline hazard and gamma distributed unobserved heterogeneity where he found the efficiency of the application of a gamma distributed heterogeneity significant. Nonetheless, to the best of author's knowledge, application of gamma distribution unobserved heterogeneity in a continuous hazard formulation has not been studied. The popularity of Weibull baseline hazard has also backed many researchers of attempting to study other parametric baseline hazards among which log-logistic is the most popular one. Therefore, there are many other proportional hazard models that either have not been properly studied in the literature or there are few applications for them.

This study aims to fill the existing gaps in the literature of parametric hazard models by providing the formulation of few less popular hazards. Two baseline hazards, namely, Weibull and log-logistic are presented and discussed in this study for four types of models: 1- continuous model without heterogeneity 2- contiguous model with heterogeneity 3- discrete model without heterogeneity and 4- discrete model with heterogeneity. Consequently, mathematical formulations for the eight types of hazard models are presented followed by an application of these models on a jointly developed household level vehicle transaction behavior.

The dataset used in this study was extracted from a survey of household automobile ownership in Toronto, Canada. The survey was conducted at the University of Toronto in 1998 (Roorda *et al.* 2000). The Database consists of information on the characteristics of more than 900 households, individual members, and their vehicles transactions and holding information. The database also includes information about residential, employment, and lifestyle changes over time for each observation. The survey covers a 9 years period from 1990 to 1998 and any other information beside this period is censored.

Model Specifications and Formulations

Starting from a simple hazard function, a continuous proportional hazard model with Weibull baseline hazard and without heterogeneity is initially presented. Cox, who pioneered the area of hazard models, in 1959 presented the early versions of hazard models with Weibull baseline hazard. The popular form of proportional hazard function is:

$$h_i(t) = \lim_{\delta \to 0^+} \frac{prob[t+\delta > T_i \geq t \mid T_i \geq t]}{\delta} = \gamma t^{\gamma} e^{-\beta x_i + \varepsilon_i} = \frac{f_i(t)}{S_i(t)} = \frac{f_i(t)}{1-F_i(t)} \qquad (1)$$

where x_i represents the covariates vector for individual *i*, parameter vector corresponding to each element of vector x_i is β, f_i is the probability of failure at time t and S_i is the probability of surviving until time t. The error term has an extreme value type distribution.

Therefore, the likelihood function for estimating the parameters of a vehicle transaction model with three transaction types (Acquisition, Trade and Dispose) would be:

$$L = \prod_{i=1}^{N_v} f_r(t)^{y_{ir}} \times S_r(t)^{1-y_{ir}} \times f_d(t)^{y_{id}} \times S_d(t)^{1-y_{id}} \times f_a(t)^{y_a} \times S_a(t)^{1-y_a} \tag{2}$$

In equation 2 probability density and survival functions can be easily calculated using the relations presented in equation 1. The hazard function presented in equation 1 can also be simply transformed to a continuous log-logistic hazard model with updating the baseline hazard:

$$h_i(t) = \frac{\frac{\beta}{\alpha}(\frac{t}{\alpha})^{\beta-1}}{1+(\frac{t}{\alpha})^{\beta}} e^{-\beta x_i + \varepsilon_i} \tag{3}$$

The continuous hazard functions of equations 1 and 3 can be enhanced with an unobserved heterogeneity parameter. The continuous survival and probability density functions that are used in the likelihood function of equation 2 with a Weibull baseline hazard are:

$$S_i(t) = \left[1+\sigma^2 \times t^{\gamma} \times e^{\beta x_i}\right]^{-\frac{1}{\sigma^2}}$$

$$f_i(t) = \gamma t^{\gamma-1} \times e^{\beta x_i} \left[1+\sigma^2 \times t^{\gamma} \times e^{\beta x_i}\right]^{-\frac{1}{\sigma^2+1}} \tag{4}$$

The exponential of the heterogeneity parameter was assumed to be distributed as a random variable with a mean of one and variance of σ^2 in equation 4.

Similarly, the log-logistic baseline hazard can be substituted with the Weibull function in equations 4 which results in the following equations:

$$S_i(t) = \left[1-\sigma^2 \times \log\left(\frac{1}{1+(\frac{t}{\alpha})^{\beta}}\right) \times e^{\beta x_i}\right]^{\frac{1}{\sigma^2}}$$

$$f_i(t) = \frac{\frac{\beta}{\alpha}(\frac{t}{\alpha})^{\beta-1}}{1+(\frac{t}{\alpha})^{\beta}} \times e^{\beta x_i} \left[1-\sigma^2 \times \log\left(\frac{1}{1+(\frac{t}{\alpha})^{\beta}}\right) \times e^{\beta x_i}\right]^{\frac{1}{\sigma^2+1}} \tag{5}$$

So far four continuous hazard models with different baseline hazard and the option of unobserved heterogeneity have been presented in equations 1, 3, 4 and 5. Alternatively, failures can be assumed to occur in discrete time intervals. Under this assumption the previous equations cannot be used anymore and a new set of equations should be utilized. Equations 6 and 7 show the discrete hazard functions without heterogeneity with Weibull and log-logistic baseline hazards, respectively.

$$prob(l^k < T_i \leq u^k) = e^{-(l^k)^{\gamma} e^{-\beta x_i}} - e^{-(u^k)^{\gamma} e^{-\beta x_i}} \tag{6}$$

where u^k and l^k are respectively the upper and lower bounds of the failure interval.

$$prob(l^k < T_i \le u^k) = e^{\log\left(\frac{1}{1+(\frac{l^k}{\alpha})^{\beta}}\right)e^{-\beta x_i}} - e^{\log\left(\frac{1}{1+(\frac{u^k}{\alpha})^{\beta}}\right)e^{-\beta x_i}} \quad (7)$$

Equations 6 and 7 can be enhanced by incorporating the unobserved heterogeneity. Bhat (1996) studied a discrete hazard model with Weibull baseline and gamma distribution heterogeneity and its application on shopping activity. One possible enhancement to his formulation would be to consider the non-monotonic log-logistic baseline hazard in such models. The probability density function that Bhat utilized plus the discrete hazard model with log-logsitic baseline hazard with heterogeneity are presented in the next two equations where equation 8 can be also found elsewhere (Bhat 1996).

$$prob(l^k < T_i \le u^k) = (1+\sigma^2 (l^k)^{\gamma} e^{-\beta x_i})^{-\sigma^{-2}} - (1+\sigma^2 (u^k)^{\gamma} e^{-\beta x_i})^{-\sigma^{-2}} \quad (8)$$

$$prob(l^k < T_i \le u^k) = (1-\sigma^2 \log(\frac{1}{1+(\frac{l^k}{\alpha})^{\beta}})e^{-\beta x_i})^{-\sigma^{-2}} - (1+\sigma^2 \log(\frac{1}{1+(\frac{u^k}{\alpha})^{\beta}})e^{-\beta x_i})^{-\sigma^{-2}} \quad (9)$$

So far a summary of the mathematical formulation of eight proportional hazard functions with different options are presented. Next, detailed modeling results of application of these models for a case study of vehicle transaction timing are presented.

Modeling Results

A modeling practice using the eight hazard models that were previously discussed is presented in this section. Determining a set of variables that can facilitate modeling household decision making procedure is a critical task. In this study, the majority of variables used in the model reflect the values of the variable at the time of vehicle transaction. The finalized set of variables used in the model were chosen considering those proved to be significant in household vehicle transaction behavioral modeling as evidenced in previous studies.

Characteristics of the vehicle that are considered by a person are very important as they affect the choice of a decision making unit (DMU). The effects of these latent characteristics of the vehicles for each individual are accounted for through *Vehicle Type Logsum* variable representing the maximum expected utility of each DMU with respect to vehicle's class and vintage choices. The vehicle type *Logsum* values (IVs) are estimated using a vehicle class and vintage choice model with a nested logit structure (Mohammadian and Miller 2003) using the same dataset (Roorda *et al.* 2000) used in this study.

In addition to the vehicle specific variables that were included in the *Logsum* estimates, a dummy variable representing whether the vehicle was leased or used at the time of transaction also included in the set of input variables. Furthermore, age, weight and fuel type of the vehicle were also tested in the model.

DMU attributes were shown to play an important role in explaining vehicle transaction decision. Size of the household and any changes in its size were also monitored at the time of the transaction and shown that these dynamic variables are also important in vehicle ownership decision.

The effect of parking cost on vehicle transaction was also found to be a significant factor in explaining the DMU's vehicle transaction behavior in the models estimated here. The summary statistics for the eight different models are shown in Table 1.

Table 1 Summary statistic of the eight hazard models

Variable	WCNH	WDNH	LLCNH	LLDNH	WCH	WDH	LLCH	LLDH
Dispose								
const	1.94 ***	2.70 ***	-10.15 ***	-10.29 ***	1.94 ***	2.70 ***	-2.65 ***	-3.48 ***
gamma/alpha	2.02 ***	1.86 ***	2441.20 ***	1067.01 ***	1.55 ***	1.86 ***	17.00 ***	25.76 ***
beta			1.55 ***	1.86 ***			1.63 ***	1.91 ***
sigma					0.000028	0.000085 *	-0.000340 ***	-0.000059 ***
Fleet	-0.30 ***	-0.31 ***	-0.30 ***	-0.31 ***	-0.30 ***	-0.31 ***	-0.29 ***	-0.31 ***
Dec Adult	-1.23 ***	-1.30 ***	-1.23 ***	-1.30 ***	-1.23 ***	-1.30 ***	-1.22 ***	-1.30 ***
Leased	-0.82 ***	-0.89 ***	-0.82 ***	-0.89 ***	-0.82 ***	-0.89 ***	-0.82 ***	-0.89 ***
Parking (log)	-0.19 ***	-0.21 ***	-0.19 ***	-0.21 ***	-0.19 ***	-0.21 ***	-0.18 ***	-0.21 ***
Home Downtown	0.48 ***	0.49 ***	0.48 ***	0.49 ***	0.48 ***	0.49 ***	0.48 ***	0.50 ***
Income (log)	0.34 ***	0.35 ***	0.34 ***	0.35 ***	0.34 ***	0.35 ***	0.34 ***	0.35 ***
Acquisition								
const	3.50 ***	4.18 ***	-8.72 ***	-1.60 *	3.61 ***	4.71 ***	-1.48 ***	-2.41 ***
gamma/alpha	1.82 ***	2.10 ***	838.97 ***	14.43 *	1.87 ***	2.37 ***	13.99 ***	19.75 ***
beta			1.82 ***	2.21 ***			1.91 ***	2.36 ***
sigma					0.23	0.46 ***	-0.004439	0.41 **
IV	-0.06 ***	-0.07 **	-0.06 **	-0.06 **	-0.06 **	-0.06 *	-0.06 **	-0.06 *
Home Downtown	0.36 ***	0.36 ***	0.36 ***	0.37 ***	0.40 ***	0.50 ***	0.37 ***	0.48 ***
Parking (log)	0.08 ***	0.08 **	0.08 **	0.08 **	0.09 **	0.11 **	0.08 **	0.11 **
Dec Adult	0.55 ***	0.53 ***	0.55 ***	0.56 ***	0.62 ***	0.82 ***	0.57 ***	0.78 ***
Income (log)	-0.11	-0.10	-0.11 *	-0.11 *	-0.13	-0.18 *	-0.12 *	-0.16
Trade								
const	4.09 ***	4.74 ***	-10.61 ***	-10.24 ***	4.09 ***	4.80 ***	-2.22	-4.14 ***
gamma/alpha	1.55 ***	2.26 ***	1435.96 ***	760.78 ***	2.04 ***	2.51 ***	21.07	34.12 ***
beta			2.02 ***	2.26 ***			2.10 ***	2.53 ***
sigma					0.12	0.44 ***	-0.000001 ***	0.43 ***
Used	-0.56 ***	-0.58 ***	-0.56 ***	-0.58 ***	-0.56 ***	-0.65 ***	-0.55 **	-0.64 ***
IV	-0.17 ***	-0.18 ***	-0.17 ***	-0.18 ***	-0.17 ***	-0.20 ***	-0.17	-0.20 ***
Fleet	-0.20 ***	-0.20 ***	-0.20 ***	-0.20 ***	-0.20 ***	-0.24 ***	-0.20 ***	-0.23 ***
Age (log)	1.11 ***	1.14 ***	1.11 ***	1.14 ***	1.13 **	1.49 ***	1.08	1.51 ***
Parking (log)	-0.13 ***	-0.13 ***	-0.13 ***	-0.13 ***	-0.13 ***	-0.15 ***	-0.13 *	-0.15 ***
Children	0.16 ***	0.17 ***	0.16 ***	0.17 ***	0.16 ***	0.19 ***	0.15	0.19 ***
Dec Adult	-0.51 *	-0.53 *	-0.51	-0.53 *	-0.51 *	-0.54	-0.50	-0.54
Job Suburb	0.14 *	0.14 *	0.14 *	0.14 *	0.14 *	0.14 *	0.15	0.14 *
L(c)	-1280.84	-1292.17	-1280.84	-1291.97	-1280.76	-1290.64	-1284.08	-1291.67
-2[L(c) - L(B)]	57.78	60.53	57.76	60.08	57.85	61.40	55.05	60.72

** Level of confidence greater than 80%, ** Level of confidence greater than 90%, *** Level of confidence greater than 95%*
W: Weibull, LL: Log-logistic, D: Discrete, C: Continuous, H: Heterogeneity, N: No

It can be seen from Table 1 that continuous models perform generally better than the discrete models, however a general conclusion about the preference of one model over another one varies on different case studies. Reviewing the results in Table 1 suggests that the variable for used car with a negative sign is significant in all of the columns in trade hazard. Similarly, the variable for leased car has a negative sign which means that leased cars are disposed earlier. Seniors and elderly drivers prefer to keep their vehicles and trade or dispose their vehicles less frequently, because the sign of the Log Driver's Age variable is positive. Probability of making a transaction is higher for households with more vehicles in the fleet. However, number of children in the household does not increase the probability of making a transaction. Households with residences in downtown area prefer to maintain their existing household vehicle composition and do not sell or acquire another one frequently. Furthermore, households working in suburb neighborhoods are also reluctant to trading a vehicle. Parking cost is an important variable which is present in all three transaction types affecting the household transaction decision with a reasonable sign. The parameters of the Vehicle Type Logsum in both trade and purchase hazards have values between 0 and 1. This is in line with the discrete choice formulation as the parameters of Logsum in a nested logit model should take values between 0 and 1 to be consistent with nested logit model derivation. While, this variable is incorporated in a duration model as an independent variable, the fact that the estimated parameters are still between 0 and 1 is interesting and can propose that integrating discrete choice and hazard-based models are possible and meaningful. Vehicle Type Logsum parameters are introduced in trade and acquisition hazards with negative sign which mean that the higher the expected utility of a vehicle is the greater would be its probability of being purchased.

Households are willing to purchase a car when the household income increases while they are disinclined to do so when their income decreases as shown in Table 1.

Conclusion

This study attempts to present a comprehensive perspective of the proportional hazard models. A typical hazard model consists of a Weibull baseline hazard and an exponential component incorporating the covariates. In this typical hazard formulation, it is assumed that failures are occurred on a continuous time scale while alternatively, a discrete assumption is occasionally employed. Beyond the commonly utilized hazard formulation, this study presented a summary of the methodology and application of several other proportional hazard models with various specifications such as continuous vs. discrete formulation, non-monotonic vs. monotonic baseline hazard and heterogeneity vs. no heterogeneity.

Having several proportional hazard functions with different capabilities in hand enables of the analyst to explore the best fit to the data. The application of the eight parametric hazard formulations on the TACUS (Roorda *et al.* 2000) retrospective dataset showed that the continuous hazard model formulation provides a better model fit than the discrete rival formulation. Nonetheless, no meaningful difference was found between the application of the non-monotonic baseline hazard and the existence of a gamma distributed unobserved heterogeneity in the case study of this paper. Therefore, further studies are needed to better examine the efficiency of different hazard formulations.

Further improvements to the modeling framework presented in this paper include applying other modeling options in hazard duration framework such as Accelerated Failure-Time models (AFT), Generalized Accelerated Failure-Time models (GAFT) and Mixed Proportional Hazard models (MPH) that can potentially improve the general model fit. Furthermore, combination of the specifications presented in this study and AFT, GAFT and MPH would also be an interesting research topic which can result in a list of comprehensive hazard model formulations.

Reference

1. Bhat C. R., (1996), A hazard-based duration model of shopping activity with nonparametric baseline specification and non parametric control for unobserved heterogeneity, *Transportation Research Part B: Methodological*, Vol. 30, No. 3, pp. 189-207
2. Cox, D. R., (1959), The analysis of exponentially distributed life-time with two types of failures, *Journal of Royal Statistical Society,* Vol. 21B, pp. 411-421
3. Han A. and Hausman J. A. (1990), Flexible parametric estimation of duration and competing risk models, *Journal of Applied Economitrics,* Vol. 5, pp. 1-28
4. Hensher D. A. and Mannering F. L., (1994), Hazard-Based Duration Models and Their Application to Transport Analysis, *Transport Reviews*, 14, pp. 63-82
5. Mohammadian A. and Rashidi T. H., (2007), Modeling Household vehicle Transaction behavior: A Competing Risk Duration Approach, *Transportation Research Record*, No. 2014, pp. 9-16
6. Mohammadian, A., and Miller E. J., (2003), Empirical Investigation of Household Vehicle Type Choice Decisions. *Transportation Research Record*: No. 1854, pp. 99-106
7. Roorda M., Mohammadian A., and Miller E.J., (2000) Toronto Area Car Ownership Study: A Retrospective Interview and Its Applications, *Transportation Research Record*, 1719, pp. 69-76
8. Yamamoto T., Kitamura R., Kimura S., (1999), A Competing Risk Duration Model of Household Vehicle Transactions", *Transportation Research Record*, No. 1676, pp.116-123
9. Yamamoto T., Madre J. and Kitamura R., (2004), An Analysis of the effects of French Vehicle Inspection Program and Grant for Scrappage on Household Vehicle Transaction, *Transportation Research part B*, Vol. 38, pp. 905-926
10. Yamamoto T., T. Matsuda and R. Kitamura, (1997), An analysis of household vehicle holding duration relative to intended holding durations, *Infrastructure Review*, No. 14, pp. 799-808