CETRA 2010 17–18 May 2010, Opatija, Croatia
First International Conference on Road and Rail Infrastructure

CAR SURVIVAL IN A NATIONAL CAR FLEET: NON-PARAMETRIC AND PARAMETRIC APPROACHES APPLIED TO FRENCH DATA

Zéhir Kolli, Ariane Dupont, Laurent Hivert
French National Institute for Transport and Safety Research (INRETS)

Abstract

We aim to determine the common demographic conditions and variables that affect light cars survival in the national fleet. The essential motivation is to identify which variables determine cars life expectancy and to understand how it can be measured: is vehicle age the only one, or other determinants needs to be considered?
For this car fleet survival study, we adopted a prospect of longitudinal analysis based on panel data. Yearly French "Parc-Auto" waves were linked together from 2000 up to 2006 to make a 7-year study, with 6795 cars described. Demographic variables like car age, total mileage, car status (the variable indicating if the car is unique, main or secondary in the household) and motorization (gasoline or diesel) are used for Kaplan-Meier curves to describe car survival in the French fleet. As data concerning car scrapping or disappearing from the fleet does not exist; all observations are considered to be right censored. In order to have an overview on car survival, we consider a new approach by building and comparing Kaplan-Meier survival curves by car age and average monthly mileage. The idea is to compare car use measured by an average mileage per month (calculated for each car like a total mileage declared divided by the age in months) with car's life expectancy measured by age and to conclude for different categories of car owners about how intensively they use their car. In addition to that we choose to proceed to a parametric study in order to find the best functional shape for the car age and car mileage statistical distributions. In order to illustrate the strong correlation between car age and car mileage to explain car life duration some bivariate probability density functions (pdf) for different categories of cars are shown.
Mortality risk increases for all car categories with age and mileage. Gasoline cars have a better survival rate than diesel cars. Car status is also a major variable to explain car survival: secondary cars have better survival rates than main cars or than unique cars. This is explained by the fact that sole cars are used more extensively than main cars or than secondary cars. Bivariate pdf given by age (in months) and total mileage in months help to show that this phenomenon finds its cause in a more extensive use of diesel cars and of cars declared as to be sole in the household (versus "main" or "secondary" cars). Models were also estimated using three-parameters- Beta, Gamma, Lognormal and Weibull distributions. Adequacy parameters and QQ-plots shows that the Weibull distribution give quite good results but the Beta model is the best compared to all other models. More importantly this study show that car mileage is a main dimension to describe car survival in the fleet.

1 Introduction

A longstanding question concerns household car survival in the national fleet. Planners concerned with energy, resource consumption and environmental quality will, more and more, take into account time for a national fleet to achieve green standards, time of diffusion of a new technology in a fleet, or expected time for disappearances of the 5, 10 or 15 percents of

oldest and most polluting cars from a given fleet. With the aim of having tools and elements of answer, they must have related forecasts on car fleet. In the majority of cases, car fleet statistics are not enough accurate. In order to have precise estimation and forecasts we can proceed to a dynamic modelisation including car registrations statistics that are usually well known and estimations of car disappearing through a survival data analysis. A serious examination of the longevity of cars will also permit to have a better understanding of car ownership behavior, to obtain better estimations and assessments of GHG emissions, to feign car fleet in the absence of direct taxes (in France a direct tax called "vignette" was aborted in 2000), or can help to simulate the reaction to a small, medium or massive introduction of a new technology like electric or hybrid cars for example.
In many applications, the natural time scale of the survival process is not calendar or clock time. Mathematical research on different time scales has been carried out by many researchers: Oakes (1995) and Kordovsky and Gertsbakh (1997) look at multiple running time scales in survival data analysis. Duchesne and Lawless (2000) and Duchesne and Rosenthal (2003) describe various advances in running time models for survival data. Cox and Oakes (1984, Section 1.2, pages 3-4) pointed out that "often the 'scale' for measuring time is clock time, although other possibilities certainly arise, such as the use of operating time of a system, mileage of a car, or some measures or cumulative load encountered". The traditional variable of lifetime measurement is the age of the car. But in order to have an overview on car survival we choose to take into account car use intensity through the annual driven (i.e. declared) mileage variable or average driven (i.e. declared) mileage per month. This variable is treated besides the age of the car, the traditional variable of lifetime measure. For the non parametric approach the main idea is to compare car use (which can depend on its motorization, status, horse power, cylinder capacity, etc...) measured by an average mileage per month with car life expectancy measured by age and then to conclude for different cars' characteristics concerning how their intensity of utilization or drivers' behavior can explain their life duration. Then, we show with a parametric approach that mileage of cars is a major variable to explain car life duration.

2 Data

We used "Parc-Auto" an annual car fleet database built from annual postal surveys conducted with a panel of 10 000 French households (about one hundred questions were asked in the survey about car ownership, car characteristics, main and secondary users, previous car characteristics, car use behavior, attitudes toward automobiles, etc...). "Parc Auto" database is merged to 2006 with car and household's characteristics concerned back to 2000. Hence a panel data set from 2000 to 2006 was constructed including 6795 observed cars and 4794 household (see table 1). From this database we use some car's specificity and characteristics like total mileage and age that are also well described. We study the impact of variables like motorization (Table 2) and car status in car life duration. Table 3 shows the proportion of cars by status. Three principal categories are considered by households for their cars and it can be a sole car (1), the main car (2) or a secondary car (3) with the respective proportions given as following: 63.88, 29.45 and 6.67 percents of our database (see table 2bis).

Table 1 Parc Auto annual waves.

Year	Number of cars	Number of households
2000	8115	6368
2001	8177	6438
2002	8249	6466
2003	8256	6523
2004	8170	6291
2005	8186	7162
2006	6257	5600
Total	56315	43943
Observations used	6795	4794

Table 2 Motorization.

Motorization	Number of cars	Number of households
Gasoline cars	3418	50.60
Diesel cars	3312	48.74
LPG (Liquefied Petroleum Gas)	30	0.44
Hybrid or Electric	2	0.03
Not described	33	0.49

Table 3 Car Status in the household.

Status	Motorization by status	Number of households	Total	Percents
Sole car	Gasoline cars	2025	4341	63.88
	Diesel cars	2285		
	Others	31		
Main car	Gasoline cars	1154	2001	29.45
	Diesel cars	822		
	Others	25		
Secondary	Gasoline cars	239	453	6.67
	Diesel cars	205		
	Others	9		

3 Non parametric analysis: Kaplan-Meier estimator

Usually, a first step in the analysis of survival data is the estimation of the distribution of the survival times. In biometrics, an analysis of Kaplan-Meier is practical to calculate the empirical life expectancy of a population of individuals. In that way, if we consider a car like an individual with a "birthdate" considered as the date of construction and if each car is followed until death, the curve may be estimated simply by computing the fraction surviving at each time. In our case study, we have no information about car destructions and consequently all observations are considered as right censored. This is an important limitation but Kaplan-Meier analysis allows us to estimate survival trends over time, even when cars lifetime

drop out or are studied for different lengths of time. For each interval, survival probability is calculated as the number of car still in the fleet divided by number of cars at risk. Cars that have been scrapped, abandoned, or that have not reached the time yet are not counted as "at risk". In a systematic way, cars that are scrapped or destructed are considered censored and are not counted in the denominator. Probability of surviving to any point is estimated from cumulative probability of surviving each of the preceding time intervals (calculated as the product of preceding probabilities). The survival distribution provides the proportion of cars surviving at the start of any interval and is equal to the probability of surviving each of the preceding intervals multiplied together. If we let S(t) be the probability that an item from a given population will have a lifetime exceeding t. For a sample from this population of size N let the observed times until death of N sample members be:

$$t_1 < t_2 < t_3 < \cdots < t_n$$

Corresponding to each t_i is m_i; the number at risk just prior to time t_i (risk-set) and d_i the number of deaths at time t_i: Note that the intervals between each time typically will not be uniform. The Kaplan-Meier estimator is the nonparametric maximum likelihood estimate of

$$S_t = \prod_{t_j<t} \left(1-\theta_j\right)$$

Is a product given by:

$$\widehat{S}_t = \prod_{t_j<t} \left(1-\widehat{\theta}_j\right) = \prod_{t_j<t} \left(\frac{r_j - d_j}{r_j}\right)$$

When there is no censoring, n_i is just the number of survivors just prior to time t_i. With censoring, n_i is the number of survivors less the number of losses (censored cases). It is only those surviving cases that are still being observed (have not yet been censored) that are "at risk" of an (observed) death.

4 Results of the Non Parametric Estimation

4.1 Motorization

From 2000 to 2006 gasoline cars are, on average, 131 months old for an estimated average mileage of 115 112 kilometers. Diesel car drivers declare much more mileage for less old cars. They do 148 286 kilometers and their cars are, on average, 108 months (see table 4). Gasoline cars do monthly on average 1002km, for diesel cars it is much more with a monthly value of 1681km per month. This statistical analysis result show a substantial difference between utilization and age for gasoline and diesel cars. As a start, we try to compare car usage for each type of motorization in order to identify their differences in terms of life expectancy (see figure 1).

Table 4 Car use by motorization.

Motorization	Average mileage (std-error)	Average age (std-error)
Gasoline cars	115112 (62517)	131 (70)
Diesel cars	148286 (78174)	108 (66)
Total sample	131472 (72563)	120 (69)

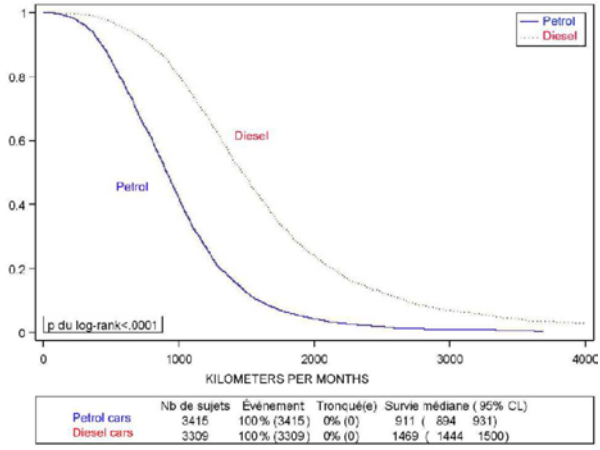


Figure 1 Kaplan-Meier curves of survival rate by motorization and per monthly average kilometers.

Figure 1 depicts results of the non parametric study. The diesel survival rate is always higher than gasoline car in term of intensity of use. Hence, diesel cars are always submitted to a more intensive use than gasoline car population. Kaplan-Meier survival analysis results reveal that fifty percents of the diesel population drive more than 1469km per month (95 percents confidence interval=1444-1500). For gasoline car it is much less with an average estimated at 911km per months (95 percents confidence interval=894-931).

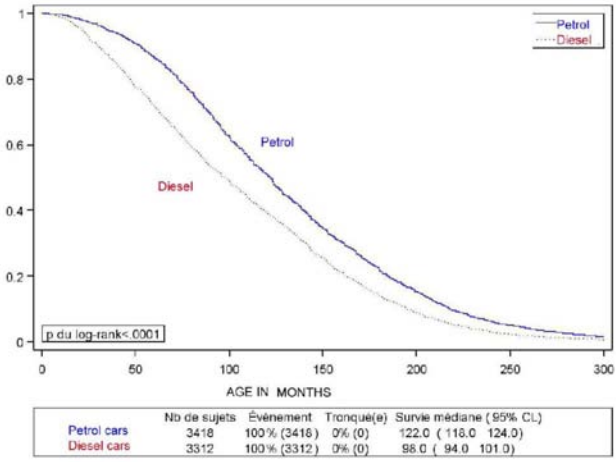


Figure 2 Kaplan-Meier curves of survival rate by motorization and per months.

The life expectancy by type of motorization is shown in figure 2. Gasoline cars have the best survival rate during the whole period. Fifty percents of gasoline cars survive after 122 months (95 percents confidence interval=118-124) it is about ten years. For diesel cars it is only 98 months, about 8 years and 2 months (95 percents confidence interval=94-101). Results are not surprising yet we know that diesel car population is submitted to a more intensive use. Diesel engine technical advantages and thermodynamic characteristics of longevity (compared to gasoline cars) are annihilated by a more intense use and their life expectancy is importantly shortened by an intensive use.

4.2 Car status

Considering motorized households only, the considerate cars can be the sole car in the household, the main car (as freely defined by respondent, usually household head) out of several cars, or a secondary car (second or more) out of several cars. This variable (sole, main, secondary) is called car status.

Table 5 Car use by status.

Status	Average mileage (std-error)	Average age (std-error)
Sole cars	124741 (70875)	104 (60)
Main cars	141162 (74056)	143 (71)
Secondary cars	153168 (72715)	168 (90)
Total	131472 (72563)	120 (69)

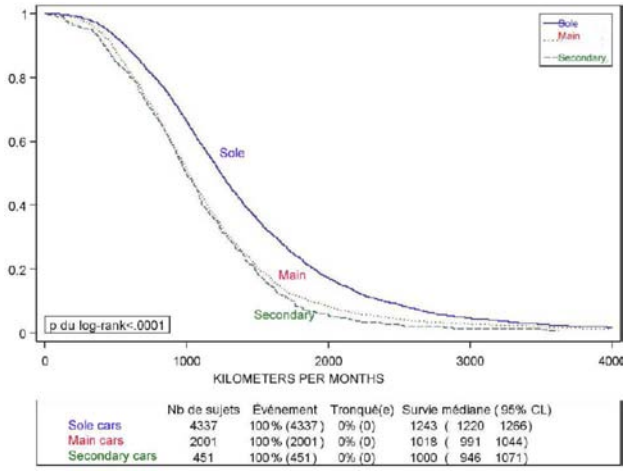


Figure 3 Kaplan-Meier curves of survival rate by car status and per monthly average kilometers.

Sole cars can be considered as the most extensively used category in the fleet comparing to other categories. Households with a unique car do in average 124741km in 104 months (see table 5). They also do in average 141162km in 143 months with their main car and 153168km in 168 months with their secondary cars (see table 5). A comparison of Kaplan-Meier curves for each category show that sole cars is far away the most extensively used category. Fifty percents of households which have a unique car do more than 1243km per months (95 percents confidence interval=1220-1246). For main car category and secondary car categories, it is almost the same level of use. With respectively fifty percents of their populations that drive more than 1018 (95 percents confidence interval=991-1044) and 1000km per months (95 percents confidence interval=946-1071).

Figure 4 depicts increasingly poorer survival with lower status category. As expected the subset of sole cars demonstrates extremely poor survival rates and the survival curve is the only one that seems to be convex. The half-population of sole cars disappear in only 95 months(nearly 8 years) (95 percents confidence interval=92-97). This is not so fast for main and secondary cars that spend respectively 141(almost 12 years) (95 percents confidence interval=137-144) and 156 months(13 years) (95 percents confidence interval=149-169) to see their half population disappear.

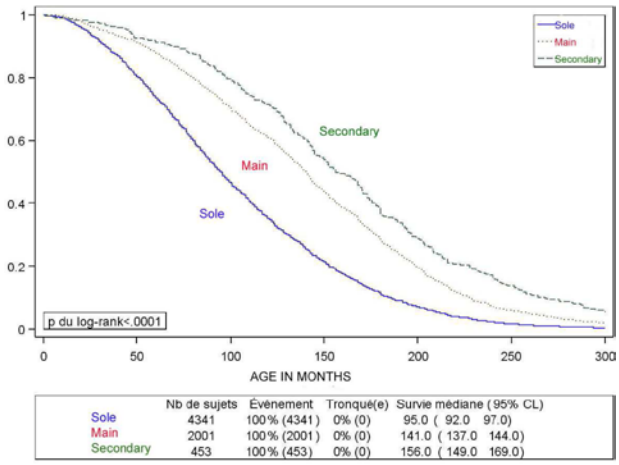


Figure 4 Kaplan-Meier curves of survival rate by car status and per months.

5 Parametric analysis

5.1 Preliminary

If we suppose that T is a continuous random variable. Then, we can define $F(t) = Pr(T < t)$ as the probability that the random variable T is less than some value t. The corresponding density function is given by $f(t) = dF(t)/dt$. It is also useful to define the survival function $S(t) = 1 - F(t) = F(t \leq T)$ that correspond to the probability that the random variable T will equal or exceed the value t. A particular useful function for duration analysis is the hazard function defined as $h(t) = f(t)/S(t)$ and gives the probability of exit from the state immediately after time t given that the state is still occupied at t. These relations shows that if either $h(t)$, $S(t)$, $f(t)$ or $F(t)$ is given, the others can be derived.

5.2 Beta distribution

The beta probability distribution function is useful for modeling random probabilities and proportions, particularly in the context of Bayesian analysis. The beta pdf has a left parameter α and a right parameter β. The beta three-parameters distribution is given by:

$$f(t,\alpha,\beta,\theta) = \begin{cases} \dfrac{(t-\theta)^{\alpha-1}\left(1-(t-\theta)\right)^{\beta-1}}{\int_0^1 (u-\theta)^{\alpha-1}\left(1-(u-\theta)\right)^{\beta-1} dt} & \text{if } t \geq 0 \\ 0 & \text{if } t < 0 \end{cases}$$

where α and β, respectively right and left parameters are positive.

5.3 Gamma distribution

The gamma three-parameters pdf of a gamma-distributed random variable *t* is given by:

$$f(t,\alpha,\beta,\theta)=\begin{cases}(t-\theta)^{\alpha-1}\beta^{-\alpha}\dfrac{e^{\frac{t-\theta}{\beta}}}{\overline{|(\alpha)}} & \text{if } t\geq 0\\ 0 & \text{if } t<0\end{cases}$$

With $\alpha \geq 0$ and $\beta \geq 0$ and $\overline{|(\alpha)} = (\alpha - 1)!$
α: shape parameter β: scale parameter and θ the location parameter ($\theta = 0$ yields to the two-parameter Gamma distribution).

5.4 Lognormal distribution

The random variable *t* follows the lognormal distribution if and only if $T = Ln(t)$ follows a standard normal distribution $N(\alpha; \beta)$. Thus, the probability density function for the three-parameters lognormal distribution is:

$$f(t,\alpha,\beta,\theta)=\begin{cases}\dfrac{1}{t.\alpha\sqrt{2\pi}}e^{-\frac{[\mathrm{Log}(t-\theta)-\beta]^2}{2\alpha^2}} & \text{if } t\geq 0\\ 0 & \text{if } t<0\end{cases}$$

5.5 Weibull and exponential distribution

The Weibull distribution has been found to provide a reasonable model for lifetimes of several types of units. The three-parameters Weibull pdf is defined as:

$$f(t,\alpha,\beta,\theta)=\begin{cases}\dfrac{\alpha}{\beta}\left(\dfrac{t-\theta}{\beta}\right)^{\alpha-1}e^{-\left(\frac{t-\theta}{\beta}\right)^{\alpha}} & \text{if } t\geq 0\\ 0 & \text{if } t<0\end{cases}$$

where α and β respectively shape and scale parameters are positive.
θ: location parameter($\theta = o$ yields to the two-parameter Weibull distribution).
Note that If $\alpha = 1$ we fall into an exponential model.

6 Modelisation

6.1 Modelisation by age

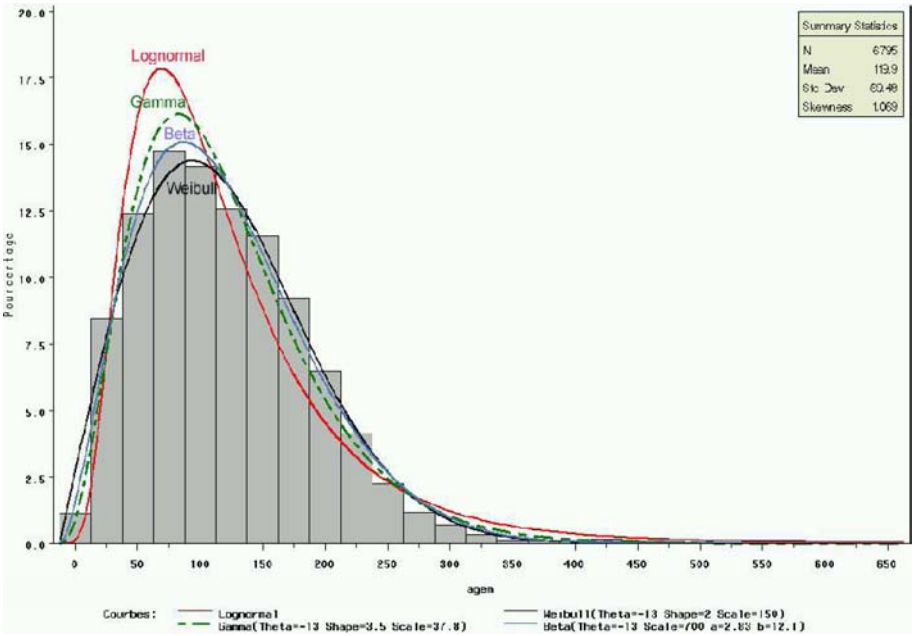


Figure 5 Functions adequacy by age

Table 6 Estimated parameters for distribution functions given by car age.

Status	Location (θ)	Scale (β)	Shape (α)	Mean	Variance
Beta (t, θ, α, β, σ(scale))	-12.5	σ=700	α = 2.829 β = 12.074	120.39	68.84
Gamma (t, θ, α, β,)	-12.5	37.826	3.502	119.93	70.78
Lognormal (t, θ, α, β,)	-12.5	4.731	0.579	122.37	85.17
Weibull (t, θ, α, β,)	-12.5	149.731	2.011	120.18	69.01

The pdf given by age in month is particularly well fitted by a Beta distribution (see Annex A for QQ-plot tests and Annex B tables 8 and 9 for quantile distributions given for Beta and Weibull distributions) although the Weibull distribution seems to give quite good results. The lognormal distribution don't provide a good fit of our sample. It has been pointed out that the hazard function for the lognormal model decreases for longer values of cycles. This does not agree with our physical understanding of progressive deterioration resulting from fatigue process.

6.2 Modelisation by mileage

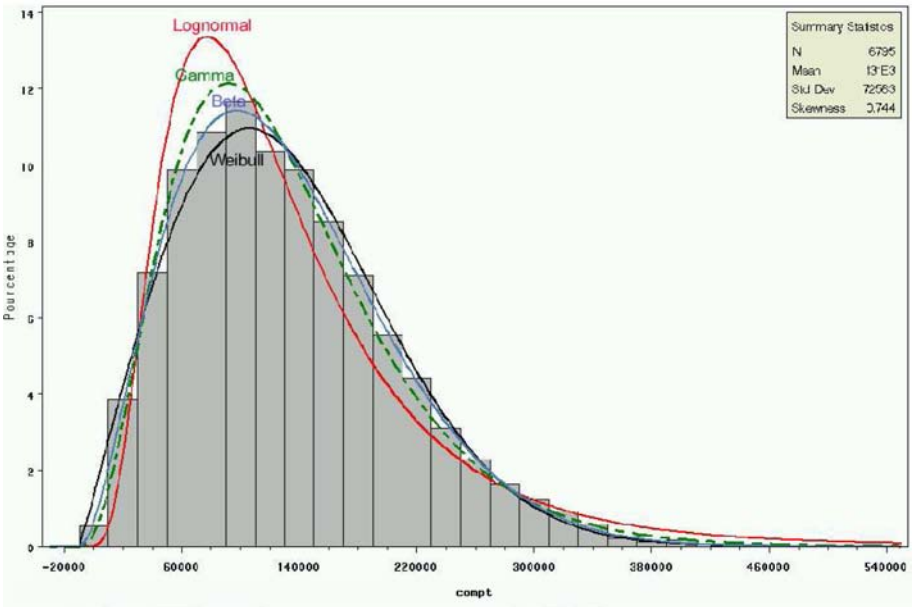


Figure 6 Functions adequacy by mileage

Table 7 Estimated parameters for distribution functions given by car mileage.

Status	Location (θ)	Scale (β)	Shape (α)	Mean	Variance
Beta (t, θ, α, β, σ(scale))	-10000	σ=700000	α = 2.876 β = 11.345	131565	72068
Gamma (t, θ, α, β,)	-10000	39896	3.546	131472	75127
Lognormal (t, θ, α, β,)	-10000	11.711	0.578	137790	105881
Weibull (t, θ, α, β,)	-10000	160053	2.064	131780	72048

The pdf given by mileage is also particularly well fitted by the Beta distribution (see Annex A for QQ-plot tests and Annex B Table 8 for quantile distributions given for Beta and Weibull distributions). The superiority of the Beta distribution is clear against the lognormal and the Gamma. But like previously, the Weibull distribution is not far of these results and from the empirical quantile distribution (Annex B Table 8). More importantly this distribution given by mileage shows the interest of a comparison by total driven mileage versus the classic comparison by age. We can also use mileage as a determinant of car life duration. More precisely, if we consider that car age is a parent process car mileage can be called marker process defined by E.A. Peña (2006): "*A marker process refers to an external process that covaries with the parent process' and hence, It assists in tracking progress of the parent process if the parent process is latent or only infrequently observed. Marker processes may also be in scientific interest in their own right. As markers of the parent process, they offer potential insights into the causal forces that are generating the movements of the parent process*". The basic analytical framework for a marker process conceives of a bivariate stochastic process X(r),Y(r) where the parent process X(r) would be car age and the marker process Y(r) the mileage (Both are one

dimensional and defined in the running time scale r). In the next section we work on these bivariate distributions to catch the link between our marker and parent process.

7 Bivariate Distributions

In the following, figure (a) to (e) represents bivariate distributions given by age and by mileage. If we let $x_i = (x_i, y_i), i \in [\![1,n]\!]$ be a sample of size *n* drawn from this distribution. The kernel density estimate of *f*(*x*,*y*) based on this sample will be given by:

$$f(x,y) = \frac{1}{n}\sum_{i=1}^{n}\varphi(x-X_i, y-Y_i) = \frac{1}{nh_xh_y}\sum_{i=1}^{n}\varphi\left(\frac{x-X_i}{h_x}, \frac{y-Y_i}{h_y}\right)$$

where $(x,y) \in \Re^2, h_x > 0, h_y > 0$ are the bandwidth and $\varphi_h(x,y)$ the rescaled normal density:

$$\varphi_h(x,y) = \frac{1}{h_xh_y}\varphi\left(\frac{x}{h_x}, \frac{y}{h_y}\right)$$

Where $\varphi_h(x,y)$ is the standard normal density function:

$$\varphi_h(x,y) = \frac{1}{2\pi}e^{\left(-\frac{x^2+y^2}{2}\right)}$$

7.1 Bivariate distributions by motorization

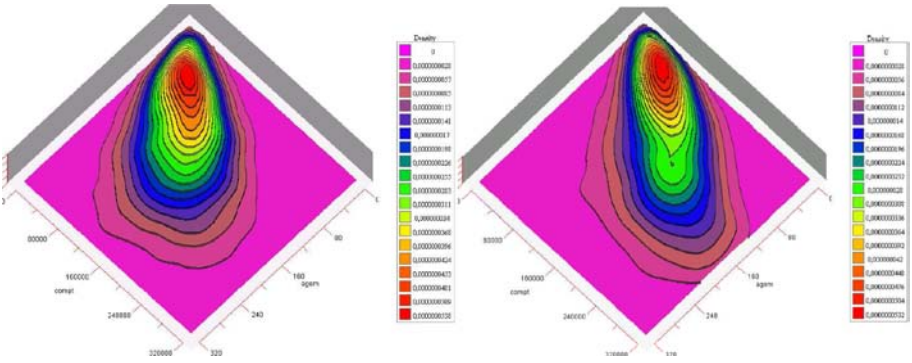
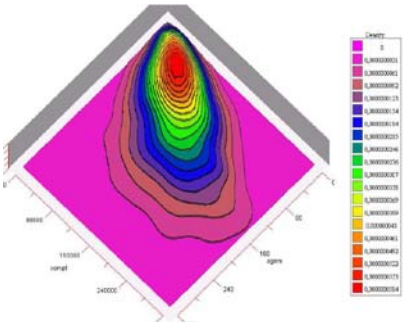


Figure 7 (a) Bivariate distribution for gasoline cars (b) Bivariate distribution for diesel cars

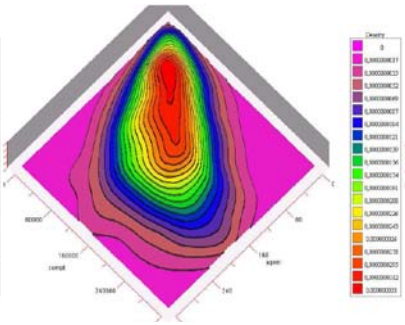
If we have a look to figure 7 (a) and (b) we can see that for the gasoline car population the bivariate pdf is centered compared to diesel car one. Gasoline car population reach its maximum for 81 months and 76 520km as for diesel cars it is around 54 months and 87 625km. The spread is also more important for diesel than for gasoline car population. In addition to that a substantial proportion of the diesel car population still in the fleet after 240 000 kilometers (some after 320 000km) while for gasoline cars the bivariate pdf ends around 260 000 kilometers.
Figures (a) and (b) enlightens us on a heterogeneousness of the behavior in automobile usage depending on motorization. Bivariate distributions looks equilibrated both by marker process which is car mileage Y(r) and parent process X(r) which is car age

7.2 Bivariate distributions by car status

The bivariate pdf is given by category of status in Figures 8, (c) (d) and (e) is enough to see that the spreading of the density increase in a radical way with the status. An other phenomenon to mention is that the age and mileage of maximums and dispersion of points increases with status. We can say that secondary cars stay longer than main cars, and that main cars stay longer than secondary car. But we also know that a quite important proportion of cars are intended to change of status one, two or more times during their lifetime in the same or in an other household.



(c) Bivariate distribution for sole cars



(d) Bivariate distribution for main cars

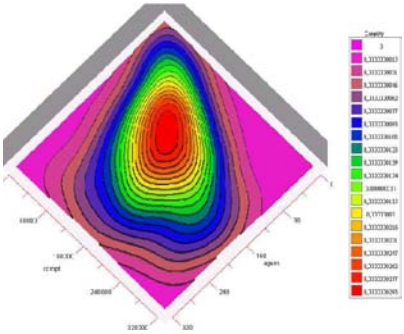


Figure 8 (e) Bivariate distribution for secondary cars.

8 Conclusion

Car status is a major variable that explain car survival in our sample. We noticed, better survival rates for cars declared as secondary cars. The reason of relatively good survival rate observed for some categories are mainly found in car use behavior. Motorization has a big impact too. Gasoline cars stay longer in the fleet than diesel cars and here too it is explained by an heterogeneity of car use behavior. Economic survival rates among these different populations should take into consideration the complex relationship among motorization and cars status of the car fleet population. To achieve the study, the estimation of pdf is of a great interest in order to derive mean or median estimations for populations or for subpopulations. Furthermore a semi-parametric approach is needed to conclude concerning the fleet structure for each population in absence of data concerning cars that are scrapped or disappear

from the fleet. In terms of motorization variables, it is clear that gasoline cars are maintained longer in the fleet than diesel cars. With the assumption that the status still the same during the whole life of each car, secondary cars are maintained longer than main cars in the French fleet. And respectively main cars are maintained longer in the fleet than sole cars. But we know that a car status is not fixed during a whole life duration and that status can move from one to another of our three categories. However, car survival duration could also be influenced by other variables: Bhat and Sen(2006), Bhat et al.(2008) conclude that vehicle-holdings and miles of travel vary with demographic characteristics, vehicle attributes, fuel costs etc... On the other hand, in our case study, car status and car motorization are quite important variables to explain and understand car life duration.

Concerning the parametric estimation, the Beta distribution often seems to provide a 'good fit' of cycles to failure data for our marker process Y(r) which is car mileage and for our parent process X(r) which is car age. It is known that, the Weibull distribution is based on more physically convincing arguments than the Beta distribution. But the Beta distribution is well suited for certain procedures of statistical extrapolation to large systems and has the advantage of dealing easily with our censored data. Car survival time can be explained both by its mileage which could be defined as a marker process or by car age defined as a classic calendar time which would be the parent process. If possible, developments of survival bivariate or multivariate function analysis should be envisaged to assess car life duration and to estimate the national car fleet size.

ACKNOLEDGMENTS: The author acknowledges the research support provided by ADEME and INRETS/DEST. He is grateful to his colleague Jean-Luc Wingert for his careful reading of the manuscript.

References

[1] Bhat, C.R. Sen, S. Eluru, N. : The Impact of Demographics, Built Environment Attributes, Vehicle Characteristics, and Gasoline Prices on Household Vehicle Holdings and Use, Transportation Research Part B (2008), 43(1): 1-18.

[2] Cox, D.R. Oakes, D. : Analysis of survival data, Chapman and Hall (1984), London.

[3] Duchesne, T. Lawless, J. P: Alternative time scales and failure time models, Lifetime Data Anal. 6 (2000),157-179.

[4] Duchesne, T. Lawless, J. P.: On the collapsibility of lifetime regression models. Alternative time scales and failure time models, Adv. in Appl. Probab.35 (2003),755-772.

[5] Peña E.A. : Dynamic Modeling and Statistical Analysis of Event Times, Statistical Science 21 (2006),487-500.

[6] Kaplan, E.L Meier, P.: Nonparametric estimation from incomplete observations, Journal of the American Statistical Association 53, (1958), 457-481.

[7] Kiefer N. M. : Economic Duration Data and Hazard Functions, Journal of Economic Litterature vol. XXVI (june 1988), 646-679.

[8] Leemis L. M. : Computing the nonparametric estimator of the survivor function when all observations are either left-or right-censored with tied observation times, Computer and Operations Research 29 (2002), 423-431.

[9] Madre J.-L., Armoogum J. : Accuracy of data and memory effects in home based surveys on travel behavior , 76th annual meeting of Transportation Research Board, Washington.

[10] Oakes, D. : Multiple time scale in survival analysis, Lifetime Data Anal. 1 (1995), 7-18.

A Annex: QQ-plots

QQ-plots are quite useful to compare data distribution with a each family of distribution that vary in location and scale. They are computed for each distribution family to compare the quantiles of the data distribution with the quantiles of the theoretical distribution family. In the following the reference line (in red) represent a particular theoretical distribution that depends on the location and scale parameters. This theoric distribution have an intercept and a slope that are equal to the location and scale parameters of that distribution.

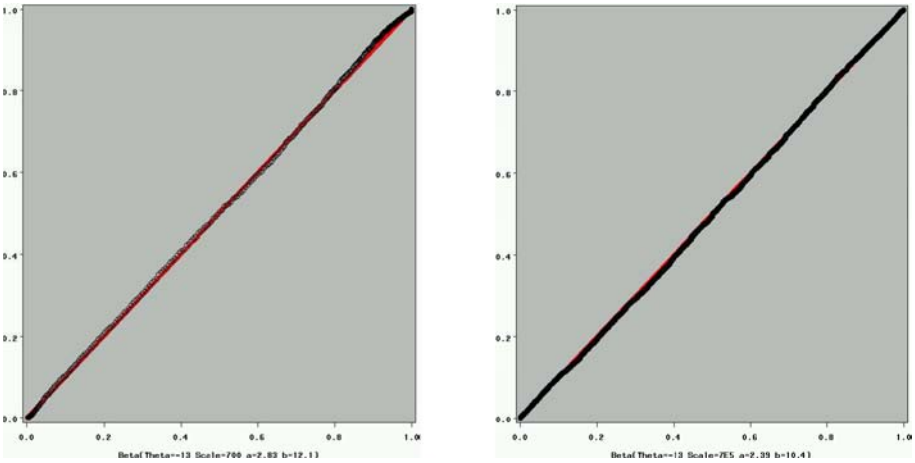


Figure 9 (h) Beta QQ-plot by AGE (i) Beta QQ-plot by Mileage

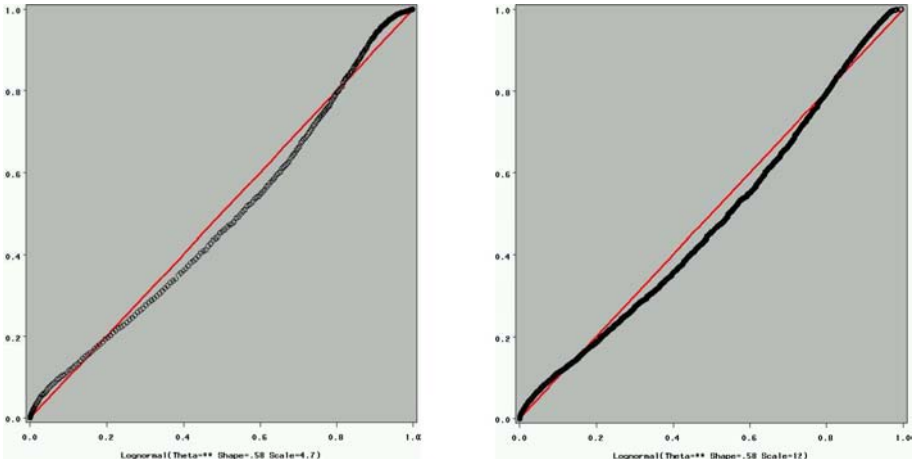
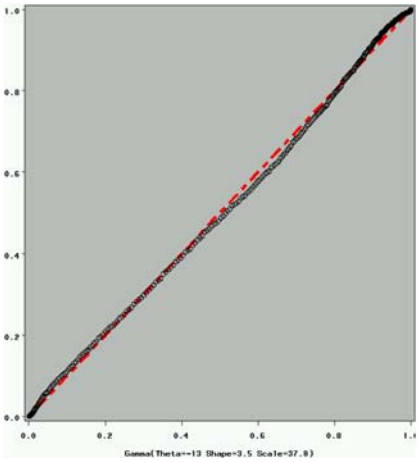


Figure 10 (j) Lognormal QQ-plot by AGE (k) Lognormal QQ-plot by Mileage



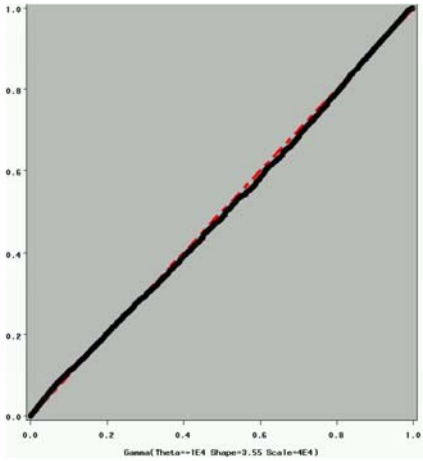
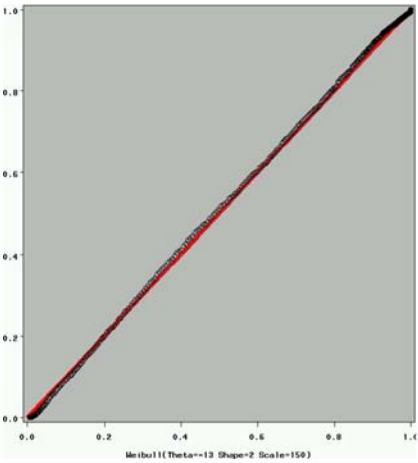


Figure 11 (l) Gamma QQ-plot by AGE (m) Gamma QQ-plot by Mileage



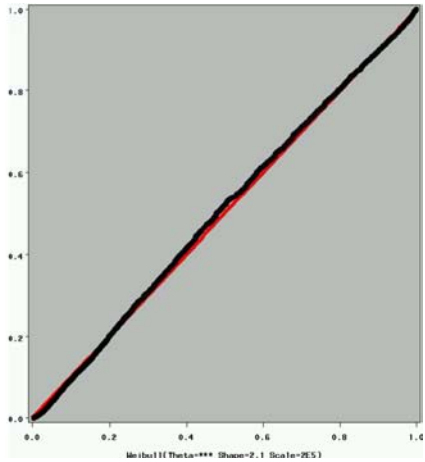


Figure 12 (n) Weibull QQ-plot by AGE (o) Weibull QQ-plot by Mileage

B Annex: Quantile distributions

Table 8 Quantile by age for observed sample, Beta and Weibull distributions.

Percents	Observed	Estimated Beta	Estimated Weibull
1.0	12	7.7	2.7
5.0	25.0	25.9	21.7
10.0	38.0	39.4	36.4
25.0	68.0	68.5	68.1
50.0	111.0	110.6	112.3
75.0	162.0	162.1	163.6
90.0	209.0	215.0	214.2
95.0	238.0	248.7	245.9
99.0	316.0	314.0	307.4

Table 9 Quantile by mileage for observed sample, Beta and Weibull distributions.

Percents	Observed	Estimated Beta	Estimated Weibull
1.0	14380	12154	7228
5.0	32000	31817	27952
10.0	45625	46265	43791
25.0	76000	77202	77515
50.0	120145	121645	124010
75.0	175059	175667	177499
90.0	230710	230627	229759
95.0	268340	265425	262365
99.0	332934	332306	325457