

PAPER • OPEN ACCESS

## Kaplan-Meier Survival Analysis in Estimating the Number of Earned Exposure Units

To cite this article: Tri Handhika and Martha Endika Sasongko 2018 *J. Phys.: Conf. Ser.* **1028** 012218

View the [article online](#) for updates and enhancements.

### You may also like

- [Sustainable development and gender well-being](#)  
Caren Grown, Maria Floro and Odera Onyechi
- [Effective Performance Evaluation to Estimate Cost and Time Using Earned Value](#)  
F Susilowati and W M Kurniaji
- [Establishing a survival prediction model for esophageal squamous cell carcinoma based on CT and histopathological images](#)  
Jinlong Wang, Lei-Lei Wu, Yunzhe Zhang et al.

## UNITED THROUGH SCIENCE & TECHNOLOGY



**248th  
ECS Meeting**  
Chicago, IL  
October 12-16, 2025  
*Hilton Chicago*



**Science +  
Technology +  
YOU!**

**Register by  
September 22  
to save \$\$**

**REGISTER NOW**

# Kaplan-Meier Survival Analysis in Estimating the Number of Earned Exposure Units

Tri Handhika<sup>1</sup> and Martha Endika Sasongko<sup>2</sup>

<sup>1</sup>Computational Mathematics Study Center, Gunadarma University, Depok, 16424, Indonesia, trihandika@staff.gunadarma.ac.id

<sup>2</sup>Master of Management, Faculty of Economic and Business, University of Indonesia, Jakarta, 10430, Indonesia

msasongko@gmail.com

**Abstract:** The estimated number of earned exposure units can help corporate executives of auto insurance company in calculating their own individual claim reserves. We conduct the estimation of this number of earned exposure units by using Kaplan-Meier survival analysis to generate the survival table. The generated survival table were then divided to be 9 risk groups based on the issuance of Financial Services Authority of Indonesia No. 6/SEOJK.05/2017 and the duration of auto insurance coverage in Indonesia. We used Root Mean Squared of Error (RMSE) as one of error measurement of the survival model. Meanwhile, a series of one-sample test with log-rank weighted function were performed for each risk group to demonstrate the goodness of fit. The estimation generated by using the survival table produce more conservative results compared to the reality. This gives assurance for one of the auto insurance company in Indonesia to use the expected earned exposure units as a basis to calculate the amount of unit that have risk to suffer a partial loss claim but still support the company in optimizing their resources.

## 1. Introduction

According to the rules issued by Financial Services Authority of Indonesia (abbreviated as OJK), the general insurance company must calculate the claim reserve. One of methods that can be used to calculate claim reserves, is namely micro-level reserving method which requires the number of earned exposure units. The capability of modern computing has enabled to process the large dataset in a short time that motivates the implementation of micro-level reserving method in industrial practice[1-3]. By obtaining an estimated number of earned exposure units over a period, it can help company executives to estimate the resource needs in the optimal claim reserve.

We need a survival model for estimating the number of earned exposure units. The usage of survival analysis was originated in clinical and health studies. Another usage is on failure times in industrial engineering, that has been evolved to several other fields[4]. One of implementation on financial fields has been introduced by Kristanti et al[5]. They have implemented survival analysis in survivability of a company caused by financial distress.

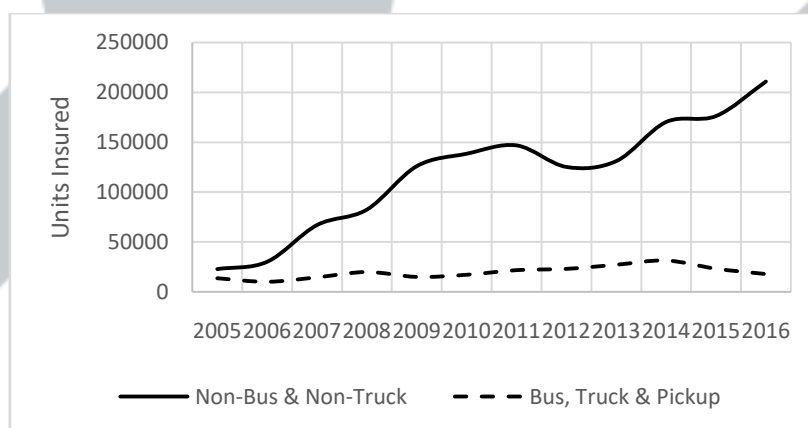
The survival model can be formed as either parametric, non-parametric or semi-parametric approach[6]. Parametric survival model based on known distribution such as normal, log normal,



exponential, gamma, Weibull and Gompertz, that needs to define of each parameter[7]. Meanwhile, the proportional hazard assumption can lead us to use semi-parametric survival model, e.g. Cox model[8]. As for non-parametric approach relies on counting process, that does not need any assumption on distribution neither proportional hazard[6]. This study focuses on the use of Kaplan-Meier survival model as one of non-parametric survival model. This is due to the difficulty of fulfilling the assumptions in the other two approaches.

## 2. Research Methodology

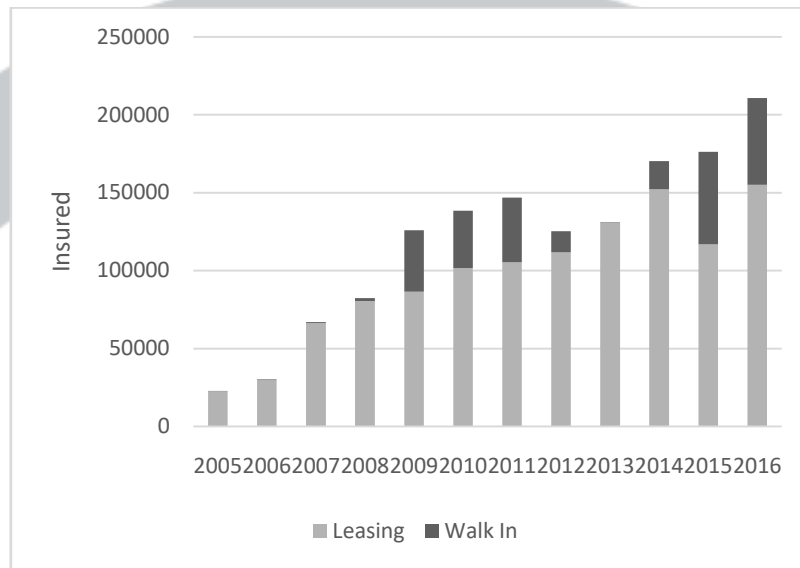
This study used incomplete data sample from one of the auto insurance company in Indonesia during the period January 1<sup>st</sup>, 2008 to December 31<sup>st</sup>, 2014. The survival model is used to estimate the number of earned exposure units between the period January 1<sup>st</sup>, 2015 and December 31<sup>st</sup>, 2016 for several risk groups. The first criteria of classification are based on the vehicle type according to the Financial Services Authority of Indonesia letter 6/SEOJK.05/2017, i.e. (i) non-bus and non-truck group, and (ii) bus, truck and pickup group. Figure 1 shows that non-bus and non-truck group has higher growth rates of units insured than bus, truck and pickup group. Hence, the risk analysis of non-bus and non-truck group has a greater impact than the other because it has larger portfolio.



**Figure 1.** The historical written number of earned exposure units for one of the auto insurance company in Indonesia.

The subsequent classification is based on the duration of coverage which are closely related to the leasing company. This is the special characteristic in the motor vehicle insurance business. It is also regulated within POJK No. 29/POJK.05/2014 Chapter V regarding the financial institution must insure each financing to mitigate the financing risk. Figure 2 shows that the written number of exposure units of non-bus and non-truck group sourced from leasing has the largest portion in each year. Therefore, in this paper, we assume that the duration of insurance coverage follows the duration of common leasing in Indonesia which is minimum 1 year and maximum 5 years. The period of coverage outside the definition was beyond the scope of the study. Finally, the risk grouping on this research consists of 9 groups: 2 groups for each 1-year, 2-year and 3-year coverage, 1 group for 4-year coverage and 3 groups for 5-year coverage. This is due to the number of different days between leap and non-leap year.

Kaplan-Meier survival model is introduced by Kaplan and Meier since 1958. This model uses a series of Maximum Likelihood Estimation (MLE) method for estimating the probability of an object aged-(x) (in days) not insured anymore in auto insurance from the beginning of the coverage year to  $x+1$  (qx). Conversely, we can estimate the probability of an object aged-(x) (in days) remaining insured in auto insurance from the beginning of the coverage year to  $x+1$  (px).<sup>9</sup>



**Figure 2.** The historical comparison of written number of earned exposure units of non-bus and non-truck group between source of businesses (leasing and walk in) for one of the auto insurance company in Indonesia.

In the Kaplan-Meier survival model, a time interval is divided into several partitions based on the time of an event in which case the auto insurance policy is terminated. The product limit estimator of  $q_x$ , is defined as:

$$\hat{q}_x = 1 - \prod_{j=1}^m (1 - \hat{q}_j) = 1 - \prod_{j=1}^m \left( \frac{r_j - d_j}{r_j} \right) \quad (1)$$

where:

- $r_j$  the number of risk in  $j$ th subinterval.
- $d_j$  the number of not insured risk anymore in  $j$ th subinterval.
- $m$  the maximum number of partitions in time interval

The product limit estimator in equation (1) is unbiased and consistent.<sup>6</sup> The product limit estimator of  $p_x$  is obtained by:

$$\hat{p}_x = \prod_{j=1}^m \hat{p}_j = \prod_{j=1}^m \left( \frac{r_j - d_j}{r_j} \right) \quad (2)$$

We use Root Mean Squared of Error (RMSE) in equation (3) as one of error measurement between  $\hat{q}_x$  and  $d_x/r_x \cdot 10$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N \left( \hat{q}_i - \frac{d_i}{n_i} \right)^2} \quad (3)$$

Furthermore, we define a random variable  $Z(\omega)$ , which is the difference between the observation  $O(\omega)$  and the expectation of the model  $E(\omega)$ . The hypothesis testing is statistically performed with the similarity between hazard function of the model and the reference hazard function. We use a central chi-squared distribution with one degree of freedom as statistical test for large sample. In order to calculate  $E(\omega)$ , we need  $Y(t)$  as the output and  $W(t)$  as a weight function characterized by  $W(t)=0$  when  $Y(t)=0$ . In this paper, we use the log-rank weighted function assuming  $W(t)=1.6$

### 3. Results and Discussion

As stated earlier, the research defines 9 groups of risk that will result survival model for each group. The Kaplan-Meier survival model in equation (2) is implemented for each duration of coverage. It will be performed more detail by non-leap and leap year separation. Figures 3a, 3b, 3c, 3d, and 3e describe



the output of Kaplan-Meier survival model for non-bus and non-truck group on 1-year, 2-year, 3-year, 4-year and 5-year coverage, respectively. We can see that there is a difference of survival probability between non-leap, 1-leap and 2-leap years. The survival probability of leap year is greater than non-leap year for each 2-year and 3-year coverage. Conversely, the survival probability of leap year is slightly less than non-leap year for 1-year coverage. Meanwhile, there is a significant difference between 1-leap and 2-leap years for 5-year coverage. The survival probability of 1-leap year is greater than 2-leap years, especially after 3 years coverage.

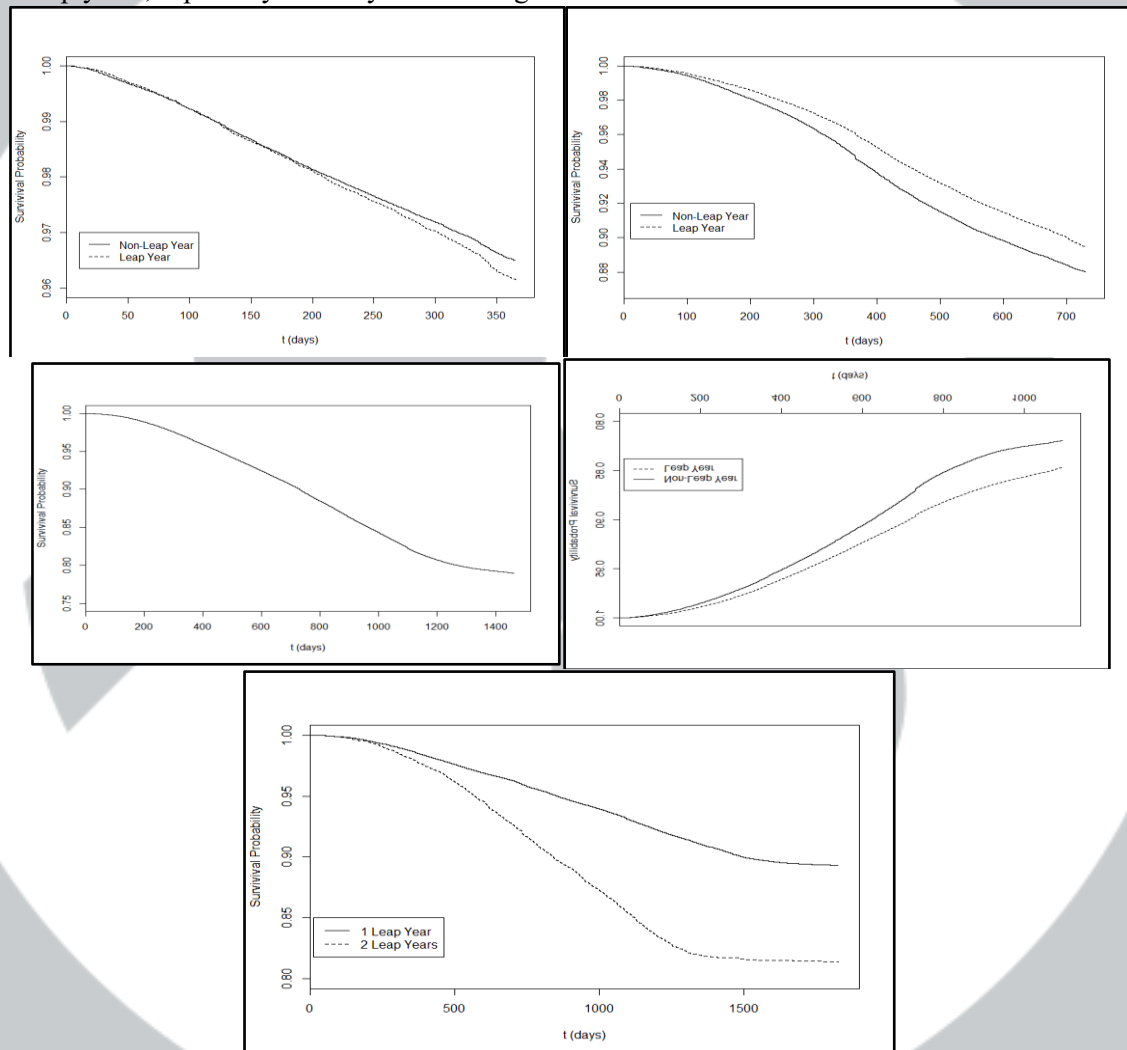


Figure 3. The Kaplan-Meier survival model of non-bus and non-truck group for one of the auto insurance company in Indonesia.

RMSE and one-sample test of Kaplan-Meier survival model for each risk groups are shown on Table 1 which is obtained by using software R 3.3.3.0. The observed data for each risk groups does not exceed critical value on the central chi-squared distribution using  $\alpha = 0.05$ . They fit a Kaplan-Meier survival model with small RMSE. It means that  $\hat{q}_x$  is close enough to  $d_x/r_x$ . Moreover, we use the Kaplan-Meier survival model resulted in Figure 3 to estimate the future number of earned exposure units between the period January 1st, 2015 and December 31st, 2016 as shown on Figure 4a, 4b, 4c, 4d, and 4e for non-bus and non-truck group on 1-year, 2-year, 3-year, 4-year and 5-year coverage, respectively.

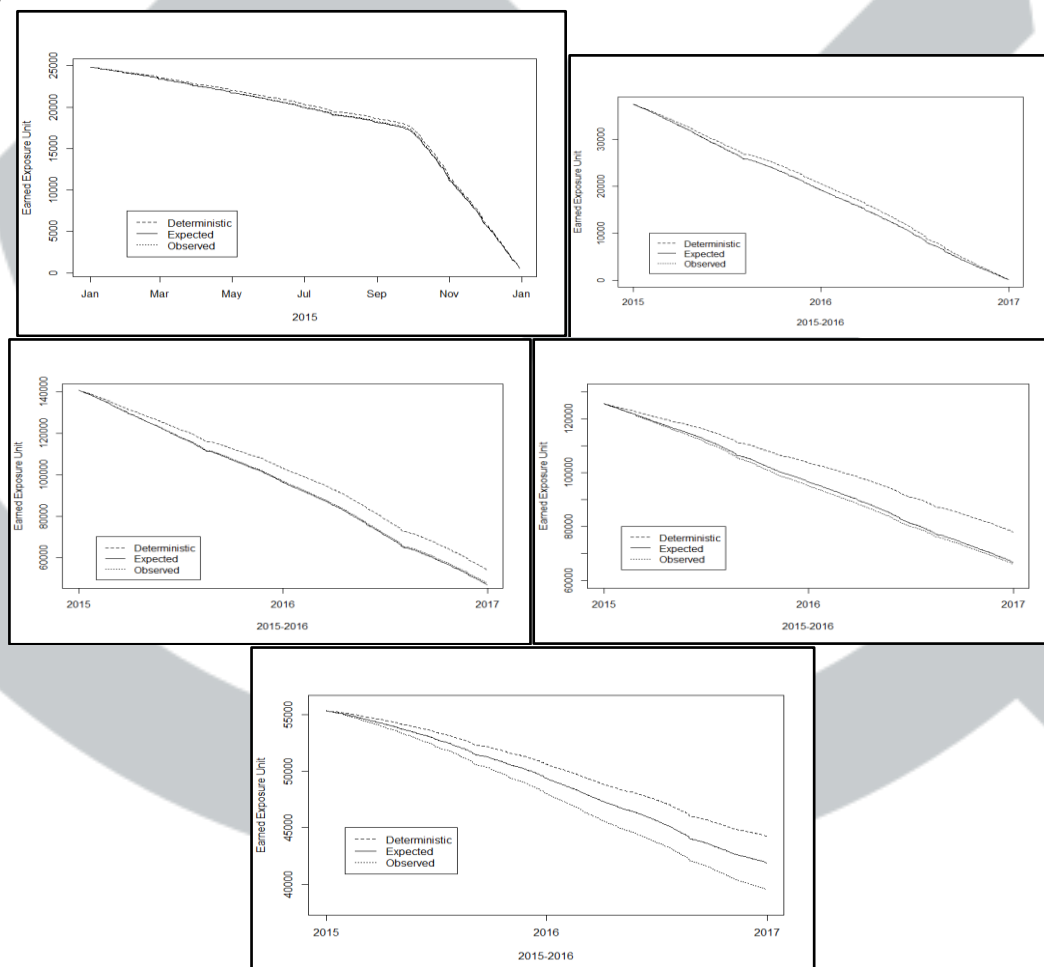
**Table 1.** RMSE and one-sample test with log-rank weighted function of Kaplan-Meier survival model for one of the auto insurance company in Indonesia.

Duration Coverage	Leap/Non-Leap	RMSE	O( $\omega$ )	E( $\omega$ )	$\chi^2_{(1)}$	p-val
1-Year	Non-Leap	6.608209	5988	5988.32	1.766	0.997
	Leap	6.756354	2340	2340.16	1.031	0.997
2-Year	Non-Leap	6.435622	9133	9134.06	12.33	0.991
	Leap	6.519708	8637	8637.84	8.213	0.993
3-Year	Non-Leap	6.793899	14911	14912.9	24.81	0.993
	Leap	6.646318	41837	41841	38.05	0.984
4-Year	Leap	7.074215	46977	46981.7	47.74	0.983
5-Year	1-Leap	6.610095	6076	6076.29	1.442	0.997
	2-Leap	5.849041	1544	1544.24	3.825	0.995

Source: This output is obtained by using software R 3.3.3.0

Notes: RMSE is multiplied by  $10^{-17}$ ,  $\chi^2_{(1)}$  is multiplied by  $10^{-5}$

We can see that the Kaplan-Meier survival model follows the real number of exposure units, especially in the shorter period. It gives more conservative estimation result than the deterministic method.

**Figure 4.** The number of earned exposure units estimation of non-bus and non-truck group for one of the auto insurance company in Indonesia.

#### 4. Conclusion

We have successfully conduct the estimation of the number of earned exposure units of non-bus and non-truck group for one of auto insurance company in Indonesia by using Kaplan-Meier survival model to generate the survival table. A series of one-sample test with log-rank weighted function and the RMSE measurement show that the observed data fit a Kaplan-Meier survival model. The estimation generated by using this survival table produce more conservative results compared to the reality. This gives assurance for one of the auto insurance company in Indonesia to use the expected earned exposure units as a basis to calculate the amount of unit that have risk to suffer a partial loss claim but still support the company in optimizing their resources.

#### References

- [1] X. Jin, *ProQuest LLC* (2014).
- [2] K. Antonio and R. Plat, *Scandinavian Actuarial Journal*, 2014, 7 (2014).
- [3] M. Pigeon, K. Antonio and M. Denuit, *Insurance: Mathematics and Economics*, 58 (2014).
- [4] J. C. Gardiner, *Statistics and Data Analysis* (2010).
- [5] F. T. Kristanti, N. Effendi, A. Herwany and E. Febrian, *Advanced Science Letters*, 22, 12 (2016).
- [6] J. P. Klein and M. L. Moeschberger, *Springer* (2003).
- [7] D. G. Kleinbaum and M. Klein, *Springer Science + Business Media, Inc.* (2005).
- [8] R. Mokarram, M. Emadi, A. H. Rad and M. J. Nooghabi, *Communications in Statistics: Simulation and Computation* (2017).
- [9] D. London, *ACTEX Publishing Inc.* (1997).
- [10] R. Jiang and W. Qian, *Open Journal of Statistics*, 3 (2013).