

Comparasion Model Analysis Time of Earthquake Occurrence in Indonesia based on Hazard Rate with Single Decrement Method

Fanny Novika¹, Ikhsan Maulidi², Budi Marsanto³, Anvika Nur Amalina⁴

^{1,3,4}Actuarial, Trisakti School of Financial Business, Indonesia

²Mathematics, Syiah Kuala University, Indonesia

novikafanny@gmail.com¹, ikhsanmaulidi@unsyiah.ac.id², bmarsanto@yahoo.com³, anvika.0107@gmail.com⁴

ABSTRACT

Article History:

Received : 31-10-2021

Revised : 27-12-2021

Accepted : 29-12-2021

Online : 22-01-2022

Keywords:

Earthquake;

Hazard Rate;

Insurance;

Single Decrement

Method;

Survival Function



The purpose of this study is to find an expectations and variability as estimators of the risk of earthquakes occurring in each province in Indonesia. Indonesia is a country that prone to natural disasters, especially earthquakes and tsunamis. The earthquake disaster damage the buildings and casualties. The risk of loss from earthquakes can be transferred using insurance. Insurance companies certainly need an analysis to estimate the probability of an earthquake occurring at a certain location and time. Hazard rate has an important role in the prediction theory of the process of earthquakes. The hazard rate can be known by the single decrement method. After the hazard rate is known, the survival function and the distribution function of the cumulative distribution of earthquake data in Indonesia will be known to look for expectations and variability as estimators of the risk of earthquakes occurring in each province in Indonesia. The data used in this study is earthquake that happen in Indonesia categorized as destructive earthquake minimum 5 magnitude. We used the data to compare a hazard function using linear model, quadratic model, cubic model and exponential. First, we plot and then using each models find the standard error. The best model suggest for Indonesia prediction Time of Earthquake Occurrence using an exponential model.



<https://doi.org/10.31764/jtam.v6i1.5535>



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license

A. INTRODUCTION

Indonesia is an archipelagic country that is prone to natural disasters, especially earthquakes and tsunamis. The main cause of this disaster was due to the tectonic conditions of the Indonesian region which is the meeting point of 3 major world plates, namely the Eurasian Continental Plate, the Indian-Australian Ocean Plate and the Pacific Ocean Plate. The meeting of these plates causes earthquakes. Active faults on land and at sea can also cause earthquakes. The occurrence of an earthquake resulted in damage to buildings and loss of life (Cipta et al., 2021).

Throughout 2016 there have been 18 destructive earthquakes in Indonesia, in 2017 there were 19 destructive earthquakes and in 2018 there were 23 destructive earthquakes (Siswadi, 2018). Among the destructive earthquakes was the Sumenep earthquake on October 10, 2018

with a magnitude of 6.4 which damaged dozens of houses and 3 people died. The biggest earthquake in the last year was the Donggala-Palu earthquake on September 28, 2018 with a magnitude of 7.5 which caused a tsunami and liquefaction of more than 2000 people died with 1000 people missing and damaged thousands of houses and the earthquake. The earthquake in Lombok occurred on August 5, 2018 with an initial magnitude of 7 followed by aftershocks. The victims in this incident were 259 people died, more than 1000 people suffered serious injuries and more than 200 thousand people suffered minor injuries and damaged various public facilities and buildings (BBC News, 2018). The damage and loss of life caused considerable financial losses. The National Disaster Management Agency (BNPb) said the total loss of the Lombok earthquake reached Rp 12.15 trillion. This figure includes damage to buildings of Rp 10.15 trillion and economic losses of Rp 2 trillion. There is a way to overcome these financial losses, namely with insurance (Gumelar, 2018).

There are a lot of models to forecasting a earthquakes, such as assessment of Point Process Models using epicentral locations data (Bray & Schoenberg, 2013). This method shows a location dominant of the earthquakes. Another method used to know a post-earthquake emergency response in a smart city using a semi-Markov model (Ghosh & Gosavi, 2017). An elaborate semi-Markov model to capture the stochastic dynamics of the events that follow an earthquake, which will be used to quantify the hazard rate to which people are exposed and estimate the restoration time. Another modelling used from exponentiated Weibull distribution to modelling a earthquake interevent time (Pasari & Dikshit, 2018). This models describe some exponentiated Weibull distribution properties, such as the survival rate, mode, median and hazard rate. Besides that, Insurance companies certainly need a numerical analysis to estimate the probability of an earthquake at a certain location and time, the expected value and the variety of the distribution of earthquakes. The hazard rate has an important role in the theory of the likelihood of an earthquake occurring. If the hazard rate is known, then the joint density distribution for the realization of earthquake data in $(0, T)$ can be known. Generally, earthquake hazard rates are estimated based on the point process likelihood equation introduced by Vere-Jones in 1995 (Daley & Vere-Jones, 2003). A lot of application that can use single decrement method. Single decrement method can be used to estimate claim occurrence in non-life insurance (Rangkuti & Sunusi, 2019). This research concluded that the level of risk in influenced by the comparison of the number of claim filed with the number of those who had accident but did not submit a claim and the ratio between time intervals of the claims occurrence from the beginning of the observation too.

Suppose X is called a random variable if S is a sample space with a probability measure and X is a real-valued function defined from members of S (Miller, 2014). The expected value is defined into two, namely:

1. If X is a discrete random variable with the probability mass function P_X , then the expected value of X is defined as $E(X) = \sum x P_X(x)$, if the above quantities converge absolutely. If the above numbers diverge, then the expected value of X is non-existent.
2. If X is a continuous random variable with probability density function f_X , then the expected value of X is defined as $E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$, if the above integral is absolutely convergent. If the above integral diverges, then the expected value of X is non-existent (Hogg, RV; McKean, JW; Craig, 2019).

The variance of X if X is a random variable can be defined as (Ghahramani, 2018).

$$\text{var}(X) = E \left[(X - E(X))^2 \right] \tag{1}$$

The cumulative distribution function $F_X(x)$ for a random variable X is the probability that X is less than or equal to a given value, denoted $F_X(x) = P(X \leq x)$. The cumulative distribution function must meet the following conditions:

1. $0 \leq F(x) \leq 1$ for every x
2. $F(x)$ is a non-descending function
3. $\lim_{x \rightarrow -\infty} F(x) = 0$ and $\lim_{x \rightarrow \infty} F(x) = 1$

The survival function $S_X(x)$ for a random variable X is the probability that X is greater than a given value, can be denoted by the formula $S_X(x) = P(X > x) = 1 - F_X(x)$. The survival function must meet the following conditions:

1. $0 \leq S_X(x) \leq 1$ for every x
2. $S_X(x)$ is a non-increasing function
3. $\lim_{x \rightarrow -\infty} S_X(x) = 1$ and $\lim_{x \rightarrow \infty} S_X(x) = 0$

The hazard function ($h_X(x)$) is the ratio of the density function and the survival function, i.e.

$$h_X(x) = \frac{f_X(x)}{S_X(x)} \tag{2}$$

In addition to the above formula, the hazard function can also be formulated with

$$h_X(x) = -\frac{S_X'(x)}{S_X(x)} \tag{3}$$

The relationship between survival function and hazard function is

$$S(b) = e^{\int_0^b h(x) dx} \tag{4}$$

Hazard function can be used to a lot of things, such as nationally consistent probabilistic tsunami hazard assessment for Indonesia. This assessment produces time-independent forecast of tsunami hazards at the coast using data from tsunami generated by local, regional and distant earthquake source (Horspool et al., 2014). Hazard can be use also to illustrate limitation of earthquake hazard mapping. But, key aspects of hazard maps often depend on poorly constrained parameters, whose values are chosen based on the mapmakers' preconceptions (Stein et al., 2012). The additional step should be done to make it more perfect, one of the way is use a single decrement. Hazard rate using the single decrement can be used to estimate the mortality in DKI Jakarta (Nisa, 2017). This The hazard rate approach using the single decrement method with the likelihood approach requires exit time information, namely the time when an earthquake occurs. Let d_{t_0} represent the number of earthquakes that occurred in the interval $(t_0, t_0 + 1]$ and $(n_{t_0} - d_{t_0})$ represent the number of earthquakes that occurred after t_0 . Because the time of occurrence for each earthquake is different, the earthquakes are considered separately and take the product of the contribution of each earthquake to the likelihood function. Likelihood L for the i -th earthquake in the

interval $(t_i, t_i + 1]$ is given by the probability density function for the occurrence of an earthquake in that interval if it is known that no earthquake occurred until t_0 . This can be stated as follows.

$$L_i = f(t_0(i)|T > t_0(i)) = \frac{f(t_0(i))}{S(t_0)} = \frac{S(t_0(i))\mu(t_0(i))}{S(t_0)}, \tag{5}$$

i.e. the i -th contribution to L. If $s_i = t_0(i) - t_0$ is the time of occurrence of the i -th earthquake in the interval $(t_0, t_0 + 1]$, with $0 < s_i \leq 1$, then

$$L_i = \frac{S(t_0+s_i)\mu(t_0+s_i)}{S(t_0)} = {}_{s_i}p_{t_0}\mu_{t_0+s_i}. \tag{6}$$

The contribution of the number of earthquakes d_{t_0} on L is $\prod_{i=1}^{d_{t_0}} {}_{s_i}p_{t_0}\mu_{t_0+s_i}$. The contribution of $n_{t_0} - d_{t_0}$ earthquakes that occur after $t_0 + 1$ is $(p_{t_0})^{n_{t_0} - d_{t_0}}$ where n_{t_0} is the number of earthquakes appears at time t_0 or later. Thus, the total likelihood of L is

$$\begin{aligned} L &= (1 - q_{t_0})^{n_{t_0} - d_{t_0}} \prod_{i=1}^{d_{t_0}} {}_{s_i}p_{t_0}\mu_{t_0+s_i} \\ &= (p_{t_0})^{n_{t_0} - d_{t_0}} \prod_{i=1}^{d_{t_0}} {}_{s_i}p_{t_0}\mu_{t_0+s_i}. \end{aligned} \tag{7}$$

To solve equation (7) for $\widehat{q_{t_0}}$, it is necessary to assume that the distribution of ${}_{s_i}p_{t_0}\mu_{t_0+s_i}$ is expressed in the form q_{t_0} .

Then we review three cases, namely if l_{t_0+s} represents the number of earthquakes after $t_0 + s$ is assumed to be linearly distributed and exponentially required to express ${}_{s_i}p_{t_0}\mu_{t_0+s_i}$.

If l_{t_0+s} represents the number of earthquakes that occur after $t_0 + s$ is a linear function, then

$$l_{t_0+s} = a + bs. \tag{8}$$

The estimator for q_{t_0} is

$$\widehat{q_{t_0}} = \frac{d_{t_0}}{n_{t_0}}. \tag{9}$$

If l_{t_0+s} states the number of earthquakes that occur after $t_0 + s$ is an exponential function, then

$$l_{t_0+s} = ab^s. \tag{10}$$

The estimator for q_{t_0} is

$$\widehat{q_{t_0}} = 1 - \exp(-\hat{\mu}) \tag{11}$$

which is the maximum likelihood estimator for q_{t_0} .

The hazard rate has an important role because the hazard rate is related to the estimated probability of an event occurring in a certain location. Based on the definition, then for the time interval $(t_0, t_0 + \Delta t_0)$, the probability of occurrence of an event is $\mu(t_0)\Delta t_0$, and the probability that there is no event in the interval $(t_0, t_0 + n)$ is

$${}_np_{t_0} = \exp\left(-\int_{t_0}^{t_0+n} \mu(y)dy\right). \tag{12}$$

This study discusses the method for estimating the earthquake hazard rate using the single decrement method. This method is adapted from the estimation method in actuarial studies which is commonly used in making mortality tables. The results of a study conducted

show that the single-decrement hazard rate estimation method is more informative than the point process hazard rate likelihood, then an estimation of temporal point process to count earthquake hazard rate in nusatenggara region(Sunusi et al., 2013).

The problem to be investigated in this study is how to find the probability of an earthquake, the expected value and the variance of the data on the occurrence of an earthquake in $(0,T)$ in each province in Indonesia. The purpose of this study is to obtain the probability of an earthquake, the expected value and the variance of the data on the occurrence of an earthquake within $(0,T)$ in each province in Indonesia by numerical analysis using the hazard rate estimation obtained from the single decrement method with the likelihood approach.

The conclusions obtained from research activities are expected to provide input to insurance companies and the government to overcome the risk of loss due to earthquakes in Indonesia in each province. The results of this study can also be used by insurance companies and the government for more information pay attention to the provinces that have the greatest chance of experiencing an earthquake disaster. The expected output of this research is journal publication so that research results can be useful for industry and society. The published research results are in the form of modeling the risk of earthquakes in Indonesia based on data in each province.

Research relevant to this research is research from Maulidi in 2014. This research is only limited to the province of Aceh, which has experienced a tsunami. In this Study, the best model to estimate time of earthquake occurrence in Aceh is cubic model. Case studies in broader research will be very useful, namely in the scope of Indonesia, therefore this research was developed. In addition, there are studies that also develop earthquake countermeasures models. The method of classifying earthquakes in Indonesia can be used Self Organizing Maps (Halim & Widodo, 2017). This method does not require special assumptions and can perform analysis on multiple variables. The result of this research is that the 4 best clusters are formed based on each of the characteristics of the number of damaged buildings, damaged roads, dead victims and the number of victims who suffer. Besides that, the grouped earthquake events based on the hypocenter using the optimization concept of the Naïve Bayes algorithm and also optimized using the adaboost algorithm which concluded that the most earthquake classifications were earthquakes with a hypocenter of 60 km to 300 km (Prathivi, 2020).

B. METHODS

The research was carried out through a data analysis process using excel 2010, IBM SPSS Statistics 25, and Minitab 21. Data used in this study based on Indonesian Agency for Meteorological, Climatological and Geophysics (Badan Meteorologi, Klimatologi dan Geofisika or BMKG) on website http://repogempa.bmkg.go.id/repo_new/. Criteria of the data was date starting at January 2010 until December 2020 geographic rectangular area is all Indonesia (6° Top Latitude- 11° Bottom Latitude and 95° Right Longitude- 141° Left Longitude). Minimum Magnitude is 5SR and Maximum 15 SR.

The method used is quantitative analysis. Quantitative research is a research approach that is required to use numbers, starting from data collection, interpretation of the data, and the appearance of the results. This study uses secondary data analysis which leads to mathematical modeling of the analyzed data. Hazard rate has an important role in forecasting

the probability of occurrence of events in a certain time interval. In this study, the hazard rate value was estimated using a single decrement method called HRSD. HRSD includes likelihood approach and moment approach. Hazard rate is usually denoted by μ_{t_0} .

Suppose $X(t_0) = T - t_0$ represents the waiting time until the occurrence of the next earthquake, if it is known that the time difference t_0 since the occurrence of the last earthquake and T is the time of reappearance between the occurrence of two earthquakes and μ, S and f represent the hazard rate, survival function, and opportunity density function, respectively. Hazard rate μ_{t_0} can be expressed as

$$\mu_{t_0} = \lim_{\Delta t_0 \rightarrow 0} \frac{P(t_0 \leq T \leq t_0 + \Delta t_0 | T > t_0)}{\Delta t_0} = \frac{f(t_0)}{S(t_0)} \tag{13}$$

By integrating (13) used $\mu_y dy = d \ln S(y)$ with a limit from t_0 to $t_0 + \Delta t_0$, it is obtained that the probability that no earthquake has occurred until $t_0 + \Delta t_0$ if it is known that there has not been an earthquake until t_0 , is

$${}_{\Delta t_0}p_{t_0} = P(T > t_0 + \Delta t_0 | T > t_0) = e^{-\int_{t_0}^{t_0 + \Delta t_0} \mu_y dy} = e^{-\int_0^{\Delta t_0} \mu_{t_0+s} ds}. \tag{14}$$

Suppose $t_0 = 0$, i.e. shortly after an earthquake occurs, we get

$${}_{\Delta t_0}p_0 = S(t_0) = P(T > t_0) = e^{-\int_0^{t_0} \mu(y) dy}. \tag{15}$$

Earthquake forecast is formulated as a conditional probability of occurrence of an earthquake up to $t_0 + \Delta t_0$, given that no earthquake has occurred until t_0 . The distribution of the reappearance time T and the waiting time for the next earthquake to occur $X(t_0)$ are respectively expressed as follows.

$$T \sim {}_{\Delta t_0}p_{t_0} \mu_{t_0} \tag{16}$$

and

$$X \sim {}_{\Delta t_0}p_{t_0} \mu_{t_0 + \Delta t_0}. \tag{17}$$

In this expression, (17) is the probability that an earthquake occurs between t_0 and $t_0 + \Delta t_0$ if it is known that no earthquake has occurred until t_0 , and

$$\int_0^{\infty} {}_{\Delta t_0}p_{t_0} \mu_{t_0 + \Delta t_0} dt = 1; \frac{d}{dt} {}_{\Delta t_0}p_{t_0} = -{}_{\Delta t_0}p_{t_0} \mu_{t_0 + \Delta t_0}. \tag{18}$$

The data used is secondary data, namely data on earthquakes that occurred in Indonesia in each of the 34 provinces in Indonesia. The data needed is the time of the earthquake, the depth of the vibrating plate and its magnitude. The recorded data are earthquakes that have the potential to damage, namely those with a magnitude above 5. The study was conducted by looking for data on earthquakes that have occurred in all provinces in Indonesia through the central BMKG institution.

The process of data analysis carried out in this study are:

1. Determine the hazard rate

The first step for data analysis is to determine the hazard rate. Hazard rate is used to estimate the probability of occurrence of an earthquake from each province. After the

probability is known, the distribution of earthquake occurrences in each province can be known.

2. Determine the survival model for the occurrence of earthquakes in each province in Indonesia

The survival model can be traced after the hazard function is found. The survival model is used to estimate the expected value and risk of an earthquake in each province in Indonesia. This study compares four models that are linear ($\mu_{t_0}^* = \beta_0 + \beta_1 t_0 + \varepsilon$), quadratic ($\mu_{t_0}^* = \beta_0 + \beta_1 t_0 + \beta_2 t_0^2 + \varepsilon$), cubic model ($\mu_{t_0}^* = \beta_0 + \beta_1 t_0 + \beta_2 t_0^2 + \beta_3 t_0^3 + \varepsilon$) and exponential that has a least standard error of the estimate and the highest R^2 .

3. Determine the expected value and variance of the distribution to estimate the risk of an earthquake in each province

The expected value of a random variable can be interpreted as a weighted average of the value of the random variable over the long term. While variance is a measure of data dispersion. The variance is a measure of the distance from the data values to the data mean. Variety can indicate the value of risk. The greater the variance, the greater the risk of the random variable to be generated. In this study, the risk of an earthquake can be estimated by looking for the variance obtained from the known survival function from the previous stage.

4. Summarizing the results of data analysis

The next step is to conclude the data that has been analyzed. The results of data analysis are in the form of the occurrence distribution function along with the expected value and variance of each province. Conclusions can be made by looking for provinces with a high risk of earthquakes. The risk limit for an earthquake is the average of all risks from 34 provinces. If the variance of each province is above the average, it is categorized as an area with a high risk of earthquakes.

C. RESULT AND DISCUSSION

To determine the probability of an event within a certain interval, the hazard rate has an important role. Maulidi in 2014 stated that this method uses two sub-methods, namely the maximum likelihood and the moment method. For example, $X(t_0) = T - t_0$ represents the waiting time for an earthquake to appear, t_0 is the time the first earthquake appears and T the next time the next earthquake appears. Let μ , S and f represent the hazard rate, survival function and probability density function. The level of danger can be expressed by

$$\begin{aligned}
 \mu_{t_0} &= \lim_{\Delta t_0 \rightarrow 0} \frac{P(t_0 \leq T \leq t_0 + \Delta t_0 | T > t_0)}{\Delta t_0} \\
 &\approx \frac{f(t_0)}{S(t_0)} \\
 &= -\frac{S'(t_0)}{S(t_0)} \\
 \mu_{t_0} &= -\frac{d \ln(S(t_0))}{dt_0} \rightarrow \mu_{t_0} dt_0 \\
 &= -d \ln(S(t_0)) \rightarrow S(t_0) \\
 &= \exp \left[-\int_{t_0}^{t_0 + \Delta t_0} \mu_y dy \right].
 \end{aligned} \tag{19}$$

To determine the likelihood estimation, exit time information is needed (the time when an earthquake occurs). Suppose d_{t_0} represents the number of earthquakes that occurred in the interval $(t_0, t_0 + 1]$ and $n_{t_0} - d_{t_0}$ represents the number of earthquakes that occurred after $t_0 + 1$. Likelihood for the i -th earthquake in the interval $(t_1, t_1 + 1]$ is

$$L_i = \frac{f(t_0(i))}{S(t_0)} = \frac{S(t_0(i)) \cdot \mu(t_0(i))}{S(t_0(i))} \tag{20}$$

If $y_i = t_0(i) - t_0$ is the time of occurrence of the i -th earthquake in the interval $(t_0, t_0 + 1]$ with $0 < y_i \leq 1$, then

$$L_i = \frac{S(t_0+y_i) \cdot \mu(t_0+y_i)}{S(t_0)} = y_i p_{t_0} \cdot \mu_{t_0+y_i} \tag{21}$$

The number of earthquakes d_{t_0} at L is $\prod_{i=1}^{d_{t_0}} y_i p_{t_0} \cdot \mu_{t_0+y_i}$ so the total likelihood is

$$L = (1 - q_{t_0})^{n_{t_0} - d_{t_0}} \cdot \prod_{i=1}^{d_{t_0}} y_i p_{t_0} \cdot \mu_{t_0+y_i} \tag{22}$$

To complete the likelihood estimation, it is necessary to assume that the distribution of $y_i p_{t_0} \cdot \mu_{t_0+y_i}$ can be expressed in the form q_{t_0} . The assumption that can be used is a linear distribution, which is in the form of

$$l_{t_0+y} = a + by. \tag{23}$$

For the value of $y = 0$ we get $a = l_t$ while for $y = 1$ we get $a + b = l_{t_0+1} \rightarrow b = l_{t_0+1} - a = l_{t_0+1} - l_{t_0} = -d_{t_0}$, so that the equation

$$l_{t_0+y} = l_{t_0} - d_{t_0}y. \tag{24}$$

Maulidi (2014) has calculated the estimated value of the estimated hazard rate for each point, which is obtained

$$\widehat{q}_{t_0} = \frac{d_{t_0}}{n_{t_0}} \tag{25}$$

is the maximum likelihood estimation for q_{t_0} and the hazard rate value can be obtained using the equation

$$\widehat{\mu}_{t_0} = \frac{\widehat{q}_{t_0}}{1 - \widehat{q}_{t_0}}. \tag{26}$$

In estimating the hazard rate value, it can be applied to model estimation for earthquakes in Indonesia. The data category selected is only data that has the potential to damage, which is above 5.00 SR. The scatterplot for earthquake magnitude data for 10 years is shown in Figure 1 from SPSS Analysis.

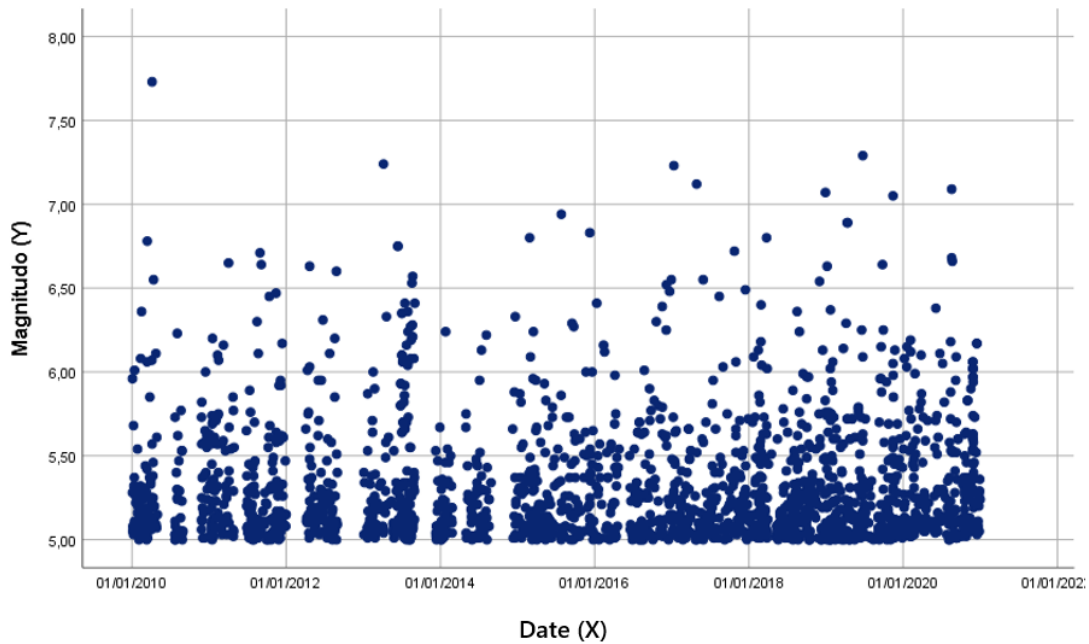


Figure 1. Plot of Magnitude against Unit of Time

The hazard rate value is estimated by using a single decrement using the likelihood estimator approach. The assumption of waiting time on the data is the assumption that the waiting time is linearly spread. The results of the calculation of the hazard rate value are in Table 1 and the hazard rate plot for each time unit is shown in Figure 2.

Table 1. Estimation Results of Single Decrement Hazard Rate with Assumption of Waiting Time Spread Linearly through Maximum Likelihood

No	Interval	Tahun	$d(t_0)$	$n(t_0)$	$q(t_0)$	$m(t_0)$
1	(0,1]	2010	140	1998	0.07007	0.07535
2	(1,2]	2011	192	1858	0.103337	0.115246
3	(2,3]	2012	85	1666	0.05102	0.053763
4	(4,5]	2013	204	1581	0.129032	0.148148
5	(5,6]	2014	107	1377	0.077705	0.084252
6	(6,7]	2015	180	1270	0.141732	0.165138
7	(7,8]	2016	145	1090	0.133028	0.153439
8	(8,9]	2017	197	945	0.208466	0.263369
9	(9,10]	2018	243	748	0.324866	0.481188
10	(10,11]	2019	266	505	0.526733	1.112971

The n_{t_0} is number of earthquakes that occurred during or after t_0 . Symbol d_{t_0} is the number of earthquakes that occur in the interval $(t_0, t_{0+1}]$. q_{t_0} is probability of an earthquake occurring at interval $(t_0, t_{0+1}]$ if it is known that there has been no earthquake until t_0 and μ_{t_0} is earthquake hazard rate immediately after t_0 .

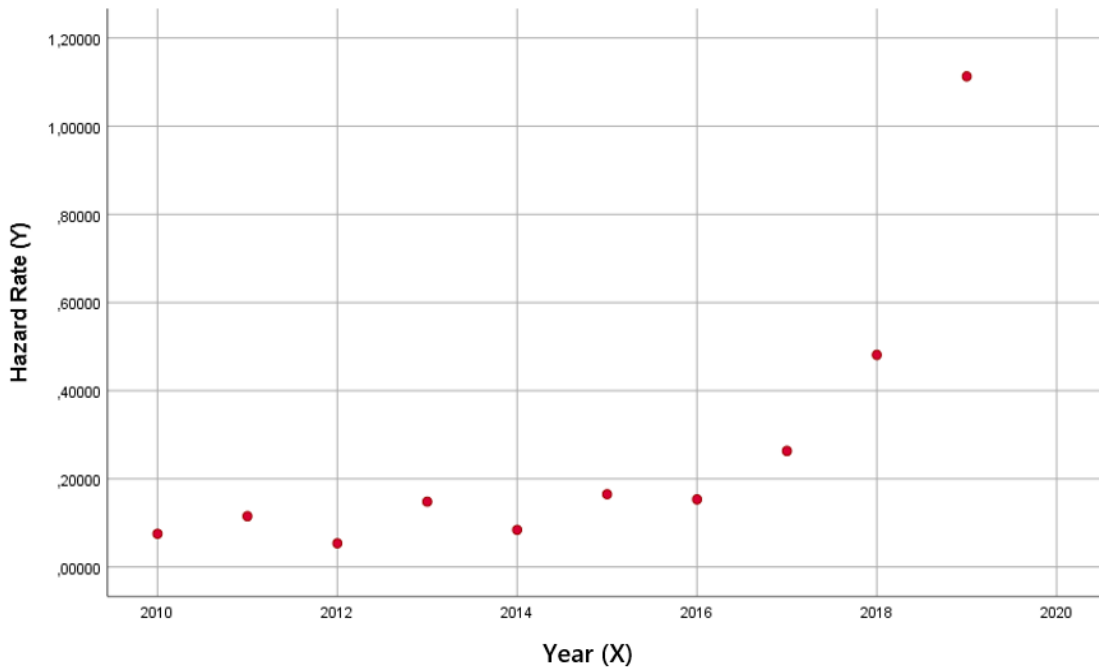


Figure 2. Plot of Hazard Rate against the Year of the Earthquake

Next, the model that fits the smallest error is determined from the actual data. To estimate the hazard function, it can be determined by estimating the coefficient with four types, namely linear ($\mu_{t_0}^* = \beta_0 + \beta_1 t_0 + \varepsilon$), quadratic ($\mu_{t_0}^* = \beta_0 + \beta_1 t_0 + \beta_2 t_0^2 + \varepsilon$), cubic model ($\mu_{t_0}^* = \beta_0 + \beta_1 t_0 + \beta_2 t_0^2 + \beta_3 t_0^3 + \varepsilon$) and exponential. Based on the analysis on SPSS, the quadratic model and the cubic model do not fit because of the collinearity based on Table 2 so that the quadratic and cubic models are not recommended to be used as estimators of the hazard rate value in earthquakes in Indonesia.

Table 2. Warning Results of Quadratic and Cubic Models on SPSS

Warnings
The Quadratic model could not be fitted due to near-collinearity among model terms.
The Cubic model could not be fitted due to near-collinearity among model terms.

1. Linear Model

Suppose $\mu(t) = a + bt$ then we get the survival function

$$\begin{aligned}
 S(t) &= \text{Exp} \left[- \int_t^{t+\Delta t} (a + by) dy \right] \\
 &= \text{Exp} \left[-ay - \frac{b}{2}y^2 \Big|_t^{t+\Delta t} \right] \\
 &= \text{Exp} \left[-at - a \cdot \Delta t - \frac{b}{2}(t + \Delta t)^2 + at + \frac{b}{2}t^2 \right] \\
 &= \text{Exp} \left[-at - a \cdot \Delta t - \frac{b}{2}t^2 - bt \cdot \Delta t - \frac{b}{2}(\Delta t)^2 + at + \frac{b}{2}t^2 \right] \\
 &= \text{Exp} \left[-\Delta t \left(a + bt + \frac{b}{2}\Delta t \right) \right]
 \end{aligned} \tag{27}$$

Let $\Delta t = c$, then

$$S(t) = \text{Exp} \left[-c \left(a + bt + \frac{b}{2}c \right) \right]. \tag{28}$$

We get the distribution function $F(t) = 1 - S(t)$,

$$F(t) = 1 - \text{Exp} \left[-c \left(a + bt + \frac{b}{2}c \right) \right]. \tag{29}$$

Then the probability density function for T is

$$\begin{aligned} f_T(t) &= F'(t) \\ &= \text{Exp} \left[-c \left(a + bt + \frac{b}{2}c \right) \right] \cdot [bc] \\ &= bc \text{Exp} \left[-c \left(a + bt + \frac{b}{2}c \right) \right] \\ &= bc \text{Exp} \left[-ac - \frac{b}{2}c^2 \right] \text{Exp}[bct]. \end{aligned} \tag{30}$$

Let $A = bc \text{Exp} \left[-ac - \frac{b}{2}c^2 \right]$, then

$$f_T(t) = Ae^{-bct}. \tag{31}$$

Expected Value μ

$$\begin{aligned} E(\mu(t)) &= \int_0^\infty (a + bt) f_T(t) dt \\ &= \int_0^\infty aAe^{-bct} dt + \int_0^\infty Abte^{-bct} dt. \end{aligned} \tag{32}$$

$$\text{(i)} \qquad \text{(ii)}$$

$$\begin{aligned} \text{(i)} \int_0^\infty aAe^{-bct} dt &= \lim_{p \rightarrow \infty} \int_0^p aAe^{-bct} dt \\ &= \lim_{p \rightarrow \infty} \left. \frac{aAe^{-bct}}{-bc} \right|_0^p \\ &= \lim_{p \rightarrow \infty} \frac{aAe^{-bcp}}{-bc} + \frac{aA}{bc}. \end{aligned} \tag{33}$$

Let $u = Abt \quad dv = e^{-bct} dt$

$$du = Ab dt \quad v = -\frac{e^{-bct}}{bc}$$

then

$$\begin{aligned} \int_0^\infty Abte^{-bct} dt &= \lim_{p \rightarrow \infty} \left. -\frac{Abte^{-bct}}{bc} \right|_0^p + \int_0^p \frac{e^{-bct}}{bc} Ab dt \\ &= \lim_{p \rightarrow \infty} \left. -\frac{Abpe^{-bcp}}{bc} + \frac{Ab0e^0}{bc} + \lim_{p \rightarrow \infty} \frac{e^{-bct}Ab}{-(bc)^2} \right|_0^p \\ &= \lim_{p \rightarrow \infty} \left. -\frac{Abp}{bce^{bcp}} + \lim_{p \rightarrow \infty} \left. -\frac{e^{-bcp}Ab}{(bc)^2} + \frac{Ab}{(bc)^2} \right|_0^p \right. \\ &= \lim_{p \rightarrow \infty} \left. -\frac{Ab}{ce^{bcp}(bc)} + 0 + \frac{Ab}{(bc)^2} \right. \\ &= 0 + 0 + \frac{Ab}{(bc)^2} \\ &= \frac{A}{bc^2} \\ &= \frac{bce^{-ac-\frac{b}{2}c^2}}{bc^2} \\ &= \frac{e^{-ac-\frac{b}{2}c^2}}{c}. \end{aligned} \tag{34}$$

Therefore

$$\begin{aligned} E(\mu(t)) &= ae^{-ac-\frac{b}{2}c^2} + \frac{e^{-ac-\frac{b}{2}c^2}}{c} \\ &= e^{-ac-\frac{b}{2}c^2} \left[a + \frac{1}{c} \right]. \end{aligned} \tag{35}$$

The following are the results of the linear model from the analysis to estimate the hazard function.

Table 3. Results of Estimating Hazard Functions for Linear Models

Coefficients						
	Unstandardized Coefficients		Standardized Coefficients		t	Sig.
	B	Std. Error	Beta			
Year (X)	.079	.025	.742		3.127	.014
(Constant)	-159.002	50.932			-3.122	.014

Based on Table 3, the estimation of the hazard function equation is $\mu_{t_0} = -159.002 + 0.079 t_0$. Meanwhile, the summary of the model conclusions is shown in Table 4.

Table 4. Summary of Suitability of Linear Model

Model Summary				
R	R Square	Adjusted R Square	Std. Error of the Estimate	
.742	.550	.494	.230	
The independent variable is Year (X).				

Table 4 are get from regression model using SPSS that provides information that 55% of the t_0 value can predict the hazard function with an estimator of an error value of 0.23. The expected value for this model is $E(\mu(t)) = e^{159.002c - \frac{0.079}{2}c^2} \left[-159.002 + \frac{1}{c}\right]$, where $c = \Delta t$

2. Exponential Model

The following are the results of the exponential model of the analysis to estimate the hazard function.

Table 5. Results of Estimating Hazard Functions for the Exponential Model

Coefficients						
	Unstandardized Coefficients		Standardized Coefficients		t	Sig.
	B	Std. Error	Beta			
Year (X)	.260	.053	.866		4.887	.001
(Constant)	2.735	.000			.	.
The dependent variable is ln(Hazard Rate (Y))						

Based on Table 5, the estimation of the hazard function equation is $\mu_{t_0} = 2.735 + 0.260 t_0$. Meanwhile, the summary of the model conclusions is shown in Table 6. The expected value for this model is $E(\mu(t)) = e^{-2.735c - \frac{0.260}{2}c^2} \left[2.735 + \frac{1}{c}\right]$, where $c = \Delta t$.

Table 6. Summary of the Suitability of the Exponential Model

Model Summary				
R	R Square	Adjusted R Square	Std. Error of the Estimate	
.866	.749	.718	.484	
The independent variable is Year (X)				

Table 6 provides information that 74.9% of the t_0 value can predict the hazard function with an estimated error value of 0.484.

D. CONCLUSION AND SUGGESTIONS

Based on the results and data analysis, it can be concluded that The estimation of earthquake modeling in Indonesia can be done using the hazard rate function using the likelihood estimator in the single decrement method. To complete the likelihood estimator, it is necessary to assume that the distribution of $y_i p_{t_0} \cdot \mu_{t_0+y_i}$ can be expressed in the form q_{t_0} . The assumption that can be used is a linear distribution, so we get $\widehat{q}_{t_0} = \frac{d_{t_0}}{n_{t_0}}$ and $\widehat{\mu}_{t_0} = \frac{\widehat{q}_{t_0}}{1-\widehat{q}_{t_0}}$. The estimation of the hazard rate function that has the best model fit value is the exponential model with $R^2 = 74.9\%$. The value of R^2 for the linear function is 55%, while the quadratic and cubic models are declared unfeasible.

ACKNOWLEDGEMENT

The authors thank the institutions (STMA Trisakti) and parties who have contributed during the research and as references.

REFERENCES

- BBC News. (2018). *Gempa Lombok: Korban Meninggal Dunia Mencapai 436 orang, Kerugian Tembus Rp5 Triliun*. Wwww.Bbc.Com. <https://www.bbc.com/indonesia/indonesia-45165817>
- Bray, A., & Schoenberg, F. P. (2013). Assessment of point process models for earthquake forecasting. *Statistical Science*, 28(4), 510–520.
- Cipta, A., Rudyanto, A., Afif, H., Robiana, R., Solikhin, A., Omang, A., & Hidayati, S. (2021). Unearthing the buried Palu–Koro Fault and the pattern of damage caused by the 2018 Sulawesi Earthquake using HVSR inversion. *Geological Society, London, Special Publications*, 501(1), 185–203.
- Daley, D. J., & Vere-Jones, D. (2003). *An introduction to the theory of point processes: volume I: elementary theory and methods*. Springer.
- Ghahramani, S. (2018). *Fundamentals of Probability: With Stochastic Processes*. Chapman and Hall/CRC.
- Ghosh, S., & Gosavi, A. (2017). A semi-Markov model for post-earthquake emergency response in a smart city. *Control Theory and Technology*, 15(1), 13–25.
- Gumelar, G. (2018). *BNPB Sebut Total Kerugian Gempa Lombok Capai Rp12 Triliun*. Cnnindonesia.Com. <https://www.cnnindonesia.com/ekonomi/20180910124912-532-329123/bnpb-sebut-total-kerugian-gempa-lombok-capai-rp12-triliun>
- Halim, N. N., & Widodo, E. (2017). Clustering Dampak Gempa Bumi di Indonesia Menggunakan Kohonen Self Organizing Maps (SOM). *Prosiding SI MaNIs (Seminar Nasional Integrasi Matematika Dan Nilai-Nilai Islami)*, 1(1), 188–194.
- Hogg, RV; McKean, JW; Craig, A. (2019). *Introduction to Mathematical Statistics Sixth Edition* (8th ed.). Prentice Hall Upper Saddle River, NJ.
- Horspool, N., Pranantyo, I., Griffin, J., Latief, H., Natawidjaja, D. H., Kongko, W., Cipta, A., Bustaman, B., Anugrah, S. D., & Thio, H. K. (2014). *A probabilistic tsunami hazard assessment for Indonesia*.
- Maulidi, I. (2014). *Pendugaan Hazard Rate Gempa di Provinsi Aceh dengan Metode Single Decrement*. IPB University.
- Miller, I. &. (2014). *Mathematical Statistics with Applications* (8th ed.). Pearson Education, Inc.
- Nisa, K. (2017). Pendugaan Hazard Rate Kematian di Provinsi DKI Jakarta dengan Metode Single Decrement Pendekatan Likelihood. *Indonesian Journal on Software Engineering*, 03(2), 24–31.
- Pasari, S., & Dikshit, O. (2018). Stochastic earthquake interevent time modeling from exponentiated Weibull distributions. *Natural Hazards*, 90(2), 823–842.
- Prathivi, R. (2020). Optimasi Algoritme Naive Bayes Untuk Klasifikasi Data Gempa Bumi di Indonesia Berdasarkan Hiposentrum. *Telematika*, 13(1), 36–43.
- Rangkuti, A., & Sunusi, N. (2019). Estimating Claim Occurrences in Non Life Insurance By Using Single Decrement Environment Method. *Journal of Physics: Conference Series*, 1341(9), 92015.
- Siswadi, A. (2018). *23 Gempa Merusak di Indonesia Sepanjang 2018*. Tekno.Tempo.Co. <https://tekno.tempo.co/read/1159731/23-gempa-merusak-di-indonesia-sepanjang-2018>

- Stein, S., Geller, R. J., & Liu, M. (2012). Why earthquake hazard maps often fail and what to do about it. *Tectonophysics*, 562, 1–25.
- Sunusi, N., Kresna, A. J., & Islamiyati, A. (2013). Hazard Rate Estimation of Temporal Point Process, Case Study: Earthquake Hazard Rate in Nusatenggara Region. *International Journal of Physical and Mathematical Sciences*, 7(6), 1059–1062.