Estimating Claim Occurrences in Non Life Insurance By Using Single Decrement Environment Method

Aidawayati Rangkuti¹, Nurtiti Sunusi²

¹Department of Mathematics, Faculty of Mathematics and Natural Sciences, Hasanuddin University. Jl. Perintis Kemerdekaan Km 10 Tamalanrea, Makassar, Indonesia
²Department of Statistics, Faculty of Mathematics and Natural Sciences, Hasanuddin University. Jl. Perintis Kemerdekaan Km 10 Tamalanrea, Makassar, Indonesia

ntitisanusi@gmail.com

Abstract. Nowadays people are increasingly aware of the importance of financial protection. In insurance companies, there are often a large number of unexpected claims. The purpose of this study is to forecast probability of claim occurrences (conditional failure rate) at certain time interval using existing historical data. To estimate the conditional failure rate, we used single decrement environment. This approach considers only one factor in the calculation process that is the time of the claim. The assumptions used for the time between two claims are exponentially and linearly distributed. The results show that both of linear and exponential assumptions, the probability of claim occurrence in time interval \(\frac{t}{t+1}\) depend on the number of claims and the number of people who had an accident at the interval of observation. Besides that, special for exponential, it depends on the ratio between time interval and the number of days at the interval of observation.

1. Introduction

Today, human needs are increasing. In this case it is seen in terms of economic needs. Erratic economic needs often cause a loss due to a problem. Someone will always try to overcome the losses that befall him. One way that someone does is financial protection by insuring it. Loss of an object, illness, and accident always lurks the lives of everyone. There is no life that is 100 per cent safe from these disorders. The interference is a risk that must be faced at any time and it is not uncommon to drain the assets and assets that we have.

Insurance is one form of risk control carried out by transferring risk from one party to another in this case an insurance company. The definition of 'risk' in insurance is "uncertainty about the occurrence of an event that can cause economic losses". The world of insurance is very synonymous with risk management. Understandably, insurance is one of the techniques in risk management. An insurance company is a company that accepts the transfer of risk from the insured. So that the company's daily activities are managing the risks of others. Most authors today tend to agree that there are important
benefits from using structurally descriptive probabilistic models in insurance [1]. Assuming that the randomness of the future accident status of individual policyholders can be described by stochastic process [9] [10]. On the other hand, the randomness can be described by a Continuous-Time Semi-Markov Inference [3] and semi-Markovian multistate model [2].

Until now, it is certain that only a handful of insurance companies formally have risk management guidelines, policies or procedures. In substance, insurance companies have carried out risk management principles, but have not been comprehensive. Some insurance companies that try to implement risk management are currently looking for forms. There is no definite guide so that the application of risk management is still groping, unlike in banking. If BI determines eight types of risks in the banking industry, but both the government and insurance associations, it has not determined the types of risks in the insurance industry [4]. The purpose of implementing risk management in the insurance industry is basically not different from other industries, namely in order to minimize and manage risks that have a negative impact on the company's goals, vision and mission [5]. In the basic theory of risk management, the stages are to determine the context (scope and purpose), risk identification, risk analysis, and risk control. Because risks are dynamic, review and monitoring must always be carried out.

To implement it, risk management guidelines are needed that can contain risk management policies and procedures. In addition, there must be an implementation so that a risk management organization structure is needed and who is involved in its implementation. For each type of company the forms can vary, both policies, procedures, organizational structures, and the people involved. In terms of structure, for example, large companies may need a special unit to handle risk management. But for other companies, risk management functions can be 'affixed' to units within the company. It is time for insurance circles to formulate risks that have the potential to disrupt the continuity of the company. Moreover, risk management is carried out by preparing contingency plans for risks that are likely to be high and have a large impact. Thus, risks that threaten the company's goals can be controlled properly. Therefore, this study aims to determine the risk of the emergence of claims on insurance companies in certain time intervals. In this study it is assumed that the time of claim submission is exponentially distributed [6].

Modeling insurance claim counts is a critical component in the rate making process for property-casualty insurers. Typically insurance companies keep a comprehensive record of the claim history of their customers and have access to an additional set of personal information. The frequency of claims reveals the risk of the insured. So by checking the relationship between the number of claims and the characteristics of policyholders, insurance companies classify policyholders and determine appropriate premiums according to the level of risk [8].

2. Materials and Methods

The failure rate is defined for non repairable populations as the (instantaneous) rate of failure for the survivors to time $t$ during the next instant of time. It is a rate per unit of time similar in meaning to reading a car speedometer at a particular instant and seeing 45 mph. We view the duration time for individual $i$, $X_i$, as a random variable having some p.d.f $f(x)$ and c.d.f $F(x)$. In characterizing the dynamics of duration
times, it is convenient to consider the hazard rate $h(x) = \frac{f(x)}{S(x)}$, which is the conditional likelihood that the event of interest occur at duration time $t$, given that it has not occurred in the duration interval $(0,t)$ [7]. The next instant the failure rate may change and the units that have already failed play no further role since only the survivor count. Hazard Rate is given by

$$h(x) = \frac{f(x)}{S(x)}$$

(1)

and the survival function by [3]:

$$S(x) = \exp\left[-\int_0^x h(u)du\right]$$

(2)

The likelihood function for $n$ independent uncensored lifetimes $x_1 \leq x_2 \leq \cdots \leq x_n$ follows as

$$L = \prod_{i=1}^n f(x_i) = \prod_{i=1}^n h(x_i)S(x_i) = = \prod_{i=1}^n h(x_i)\exp\left[-\int_0^{x_i} h(u)du\right]$$

(3)

So that the log-likelihood is

$$l = \ln L = \sum_{i=1}^n \ln h(x_i) - \sum_{i=1}^n \int_0^{x_i} h(u)du$$

(4)

3. Result

3.1. Likelihood Construction

In this study, a likelihood equation will be constructed by single decrement environment that can be used to estimate the parameters of the claim filing model. We show the mathematics of these derivations in considerable detail in Maximum Likelihood Estimation [7]. Suppose $t_i$ at is the time of filing a claim $i$, is the time of filing a claim at the time to $t + 1$, and $T$ is a random variable that states the time of filing a claim. The likelihood function for filing a claim is a function of density of opportunity for submission of claims at a certain time if it is known that there has been no claim until the time, so that for filing a claim at the time $t$ obtained

$$L_i = f(t_i|T > t) = \frac{d}{dt} F(t_i|T > t) = \frac{d}{dt} \frac{F(t_i) - F(t)}{1-F(t)} = \frac{f(t_i)}{S(t)} \frac{S(t)}{S(t)} = \frac{S(t)}{S(t)}$$

(5)
Is the contribution claim occurrence \(i^{th}\) to \(L\). If we let \(\text{s}_i = t_i - t\) be the time of time submission of claim \(i^{th}\) in time interval \((t, t + 1]\) where \(0 < \text{s}_i \leq 1\), then we have

\[
L_i = \frac{S(t + \text{s}_i) \lambda(t + \text{s}_i)}{S(t)}.
\]

(6)

Since we have \( \frac{S(t + \text{s}_i)}{S(t)} = \text{s}_i \text{P}_i \) and \( \lambda(t + \text{s}_i) = \mu_{i, \text{x}} \) then

\[
L_i = \frac{S(t + \text{s}_i) \lambda(t + \text{s}_i)}{S(t)} = \text{s}_i \text{P}_i \mu_{i, \text{x}}.
\]

(7)

where \( \mu_{i, \text{x}} \) is an opportunity for no claims to arise at intervals \((t, t + \text{s}_i]\) and \( \mu_{i, \text{s}} \) is hazard rate of claim occurrence at \(t + \text{s}_i\). Thus, the likelihood function at all time of claim submission is

\[
\prod_{i=1}^{d} \text{s}_i \text{P}_i \mu_{t + \text{s}_i}.
\]

(8)

The claim submission process in model point process, the total likelihood function can be expressed by taking into account the number of accidents but not submitting claims at a time \(t\) that is notified \((n-d)\) of the chance of no claims that occur.

\[
\text{p}_n^{\text{a} - d} = (1 - \text{q}_t)^{n-d}
\]

(9)

with \(d\) the number of claims filed at the time \(t\) and \(p\) is the number of accidents that occur at the time and is an opportunity not to appear claims at the time \(t\) that can also be denoted \(1 - \text{q}_t\), so that the total likelihood function can be stated as follows:

\[
L = (1 - \text{q}_t)^{n-d} \prod_{i=1}^{d} \text{p}_i \mu_{i, \text{x}}.
\]

(10)

### 3.2. Hazard Rate Estimation for Claim Arrival as Exponentially Distributed

Let the time for filing a claim that is exponentially distributed and if the number of accidents in the time interval between \(t\) and \(t + t_s\) is exponential function in the estimated interval between \(t\) and \(t + 1\) then the function can be formed \(L_{t + t_s} = \alpha \beta^{x_i}\), where \(x_i\) has \(0 < x_i \leq 1\) in estimation interval \([t, t + 1]\) denote by

\[
\text{s}_i = \frac{\text{number of days in estimation interval}}{x_i}.
\]

Because of the data for the claim time interval is continuous data, to ensure continuity of the data, it is assumed \(s_i = 0\) and \(s_i = 1\).

- If \(s_i = 0\) then \(L_t = \alpha\) or the number of accident at time \(t\) assumed equals \(a\).
• If \( s_i = 1 \) then \( l_{t+1} = ab \) or the number of accident in interval \( (t, t + 1] \) assumed equals \( ab \). Thus it can be expressed by \( b = \frac{l_{t+1}}{a} = \frac{l_{t+1}}{l_t} \).

So the number of accidents at intervals \( (t, t + s_i] \) can be stated as follows:

\[
l_{t+s_i} = ab^s_i = l_t \left( \frac{l_{t+1}}{l_t} \right)^{s_i} = (l_t)^{s_i} / l_t^{1-s_i}
\]  

(11)

Which is an opportunity for no claim to occur at the time interval\( (t, t + 1] \) denoted by \( p_t \). So \( l_{t+s_i} = l_t (p_t)^{s_i} \). Thus, we have that there is no claim in time interval \( (t, t + 1] \) as follows:

\[
p_t = \frac{l_{t+1}}{l_t} = (p_t)^{s_i}.
\]  

(12)

For exponential distribution, hazard rate in time interval \( (t, t + s_i] \) denoted \( \mu_{t+1} \) is constant

\[
\mu_{t+1} = \mu_{t+2} = \mu_{t+s_3} = \ldots = \mu_t = \mu,
\]

with \( \mu_{t+s_i} = \mu = \frac{d}{l_t} = -\ln p_t \). Probability that no claim in time interval \( (t, t + s_i] \) denoted by \( \mu = -\ln p_t \), or we have

\[
p_t = e^{-\mu}.
\]  

(13)

Because of event time is exponentially distribution, then \( \mu_{t+s_i} \) is constant that is \( \mu \), where \( \mu = -\ln p_t \). We know that \( p_t = e^{-\mu} \) so \( (p_t)^{s_i} = (e^{-\mu})^{s_i} \) so that

\[
q_t = e^{-\mu t}.
\]  

(14)

Thus, we have

\[
L = (1-q_t)^{n-d} \prod_{i=1}^{d} p_i \mu.
\]  

(15)

The probability of claim occurrence in time interval \( (t, t + 1] \) denoted by \( q_t = 1 - p_t \), then \( p_t = 1 - q_t \) so we have

\[
L = (1-q_t)^{n-d} \prod_{i=1}^{d} p_i \mu
\]

\[
= (p_t)^{n-d} \prod_{i=1}^{d} p_i \mu.
\]  

(16)

Because \( p_t = e^{-\mu t} \), so that

\[
L = \mu^d \exp(-\mu(n - d) - \mu \sum_{i=0} p_i s_i).
\]  

(17)

Because of (17) is a nonlinear equation, then \( l = \ln L \), we can denote \( l \) as

\[
l = d \ln \mu - \mu(n - d) - \mu \sum_{i=0} p_i s_i.
\]  

(18)

So we have
Thus, we have hazard rate as follows
\[ \frac{d}{\mu} = \left[ \frac{(n-d) + \sum s}{(n-d) + \sum s} \right] \]  
(19)

where \( d \) is the number of claim occurrence, \( n \) is number of accident at the time interval \((t, t + 1]\) during observation and \( s_i \) is the ratio between time interval of claim occurrence and number of days in observation time interval.

3.3. Hazard Rate Estimation for Claim Arrival as Linearly Distributed

Let \( l_{t+s} \), as number of accident during time interval \((t, t + s]\) is a linear function between \( t \) and \( t + 1 \), then:
\[ l_{t+s} = a + bs. \]

For \( s = 0 \) we have: \( l_t = a \) and for \( s = 1 \), then \( l_{t+1} = a + b \), so that:
\[ b = l_{t+1} - a \text{ or } b = l_{t+1} - l_t = -d_t \]

Thus, we have:
\[ l_{t+s} = l_t - s. d_t = l_t - s. (l_t - l_{t+1}) = s. l_{t+1} + (1-s). l_t \]

So that:
\[ sp_t = \frac{l_{t+s}}{l_t} = 1 - s. \frac{d_t}{l_t} = 1 - s. q_t \]  
(21)

and
\[ sq_t = s. q_t \]

We also have
\[ 1 - s. p_t = \frac{l_{t+1}}{l_{t+s}} \frac{l_{t+s}}{l_t - s. d_t} = \frac{p_t}{1-s. q_t} \]
and
\[ 1 - s. q_{t+s} = 1 - 1-s. p_{t+s} = \frac{1-s. q_t - p_t}{1-s. a_t} = \frac{(1-s). q_t}{1-s. a_t} \]

Next,
\[ \mu_{t+s} = -\frac{d_t}{l_{t+s}} \frac{d_t}{l_t} - s. d_t = q_t / 1 - s. q_t \]

when multiplied by (21), leads to the very convenient result
\[ f(s|T > t) = sp_t \mu_{t+s} = q_t \]

Where \( \mu_{t+s} \) and \( f(t|T > t) \), are not defined at \( s = 0 \) and \( s = 1 \), since \( l_{t+s} \) is not differentiable there. We know the total likelihood:
\[ L = (1 - q_t)^{n - d_t} \Pi D \frac{s \mu_{t+s}}{s p_t \mu_{t+s}} \]
Where in linear assumption the \( \prod_{D} \mu_{T+2} = \rho_{T}^{d_{t}} \)

So, we have

\[
L = (1 - q_{t})^{n_{t}-d_{t}} \prod_{D} \mu_{T+2} = (1 - q_{t})^{n_{t}-d_{t}} q_{t}
\]

Thus we define the log likelihood:

\[
l = \ln L = d_{t} \ln q_{t} + (n_{t} - d_{t}) \ln (1 - q_{t})
\]

Then

\[
\frac{dl}{d\sigma_{t}} = \frac{d_{t}}{\sigma_{t}} - \frac{n_{t} - d_{t}}{1 - \sigma_{t}} \ln (1 - \sigma_{t})
\]

is the likelihood equation which easily produces

\[
\hat{q} = \frac{d_{t}}{n_{t}}
\]

which is an Maximum Likelihood Estimator of \( q \).

4. Simulation

In this section a simulation is carried out by generating 2 groups of data, each of which is exponential and linear distribution for the time between submitting two consecutive claims. The variables used are the time of filing a claim, the number of accidents, and the number of claim. Needed, including the number of people who have an accident, the number of claims, and the number of people who have an accident but do not submit a claim. This can be used

| No | Interval | \( n \) | \( d \) | \( \sum x_{i} \) | Exponential \( \mu \) | Exponential \( q \) | Linear \( q \) |
|----|----------|------|------|-------------|-----------------|-----------------|-----------------|
| 1  | (0,1]    | 256  | 1    | 0.24        | 0.00391         | 0.0039          | 0.00391         |
| 2  | (1,2]    | 255  | 1    | 0.21        | 0.00393         | 0.00392         | 0.00392         |
| 3  | (2,3]    | 254  | 1    | 0.16        | 0.00395         | 0.00394         | 0.00394         |
| 4  | (3,4]    | 253  | 1    | 0.23        | 0.00396         | 0.00395         | 0.00395         |
| 5  | (4,5]    | 252  | 3    | 2.01        | 0.01013         | 0.01008         | 0.0119          |
| 6  | (5,6]    | 249  | 17   | 8.19        | 0.07077         | 0.06832         | 0.06827         |
| 7  | (6,7]    | 232  | 19   | 9.6         | 0.08535         | 0.08181         | 0.0819          |
| 8  | (7,8]    | 213  | 24   | 12.04       | 0.11937         | 0.11252         | 0.11268         |
| 9  | (8,9]    | 189  | 8    | 2.62        | 0.04356         | 0.04262         | 0.04233         |
| 10 | (9,10]   | 181  | 3    | 1.24        | 0.01673         | 0.01659         | 0.01657         |
| 11 | (10,11]  | 178  | 4    | 3.01        | 0.02259         | 0.02234         | 0.02247         |
| 12 | (11,12]  | 174  | 3    | 0.83        | 0.01745         | 0.0173          | 0.01724         |
Table 1 shows that the number of accidents in the observation time interval are 256. For interval (0,1] that is January 1, 2010 until February, 2010 there is 1 filling claim. The probability of claim occurrence in interval (0,1] is 0.003. Furthermore, the determination of the level of risk in the interval (1,2], namely on March 1, 2010 to April 30, 2010, is similar to determining the level of risk in the previous interval, but the number of accidents that occur at that time is the number of accidents that occur at the interval observed minus claims that occurred at the previous interval, so the risk level at the interval (1,2] is 0.003. The same is done for intervals (2,3] and so on. The relationship between each interval with the probability of claim emergence is shown in the following Figure 1.

![Figure 1. Probability of claim occurrence for Exponential and Linear Distributions.](image)

In Figure 1 it can be seen that the risk level of each interval varies greatly even though some intervals have the same number of claims. Furthermore, Figure 6 shows that the probability of occurrence of claims during the observation time interval for the two models has almost the same pattern, but the linear model shows that the probability of occurrence of claims at certain intervals is higher than the exponential model.

5. Conclusion

In this study, we have carried out a study of the probability of occurrence of claims at each particular period in the interval of observation by using single decrement method. It can be concluded that the level of risk is influenced by the comparison of the number of claims filed with the number of those who had an accident but did not submit a claim. Besides that the level of risk is also influenced by the ratio between the time intervals of the claims occurrence from the beginning of the observation with the number of days at the interval of observation.

6. Acknowledgement

The authors express their sincere thanks to the Editor and a referee for their constructive criticisms and excellent suggestions which led to a considerable improvement of this paper. This research supported by Ministry of Research, Technology and Higher Education of the Republic of Indonesia through Decentralization Research of Hasanuddin University.
References

[1] Arjas E. The claims reserving problem In non-life insurance: Some structural ideas. *Astin Bulletin*. 1998, 19(2): pp139–52. doi: 10.2143/AST.19.2.2014905

[2] Adékambi, F. and M. C. Christiansen. Integral and Differential Equations for the Moments of Multistate Models in Health Insurance, *Scandinavian Actuarial Journal*, 2017(1): pp 29–50.

[3] Biessy, G. Continuous-Time Semi-Markov Inference of Biometric Laws Associated With a Long-Term Care Insurance Portfolio, *ASTIN Bulletin*, 2017, 47(02): pp 527–561.

[4] Brown, J. and M. Warshawsky. The Life Care Annuity: A New Empirical Examination of an Insurance Innovation that Addresses Problems in the Markets for Life Annuities and Long-Term Care Insurance, *Journal of Risk and Insurance*, 2013, 80(3): pp 677–703.

[5] Czado, C. and F. Rudolph. Application of Survival Analysis Methods to Long-Term Care Insurance, *Insurance: Mathematics and Economics*, 2002, 31(3): pp 395–413.

[6] Eugster, B., R. Lalive, A. Steinhauer, and J. Zweim’uller. The Demand for Social Insurance: Does Culture Matter?, *The Economic Journal*, 121(556): pp 413–448. 2011

[7] Kristiaan Helsen, David C. Schmittlein. Analyzing Duration Times in Marketing: Evidence for the Effectiveness of Hazard Rate Models. *Marketing Science*, 1993,12(4): pp 395-414. doi:10.1287/mksc.12.4.395.

[8] Shi, P. and Valdez, E.A. Multivariate negative binomial models for insurance claim counts. *Insurance: Mathematics and Economics*, 2014, 55, pp 18–29.

[9] Sunusi N, E.T. Herdiani, F. Nawawi and Giaro. Exploratory Analysis of Rainfall Occurrence in South Sulawesi Region Using Spatial Point Process. *J. Math. Statist*, 2015, 11: 113-118.doi: 10.3844/jmssp. pp113-118.

[10] Sunusi N, Aidawayati R, Irmayani. Study of Insurance Claim Using Point Process Models. *Indian Journal of Science and Technology*, 2016, 9(28).doi:10.17485/ijst/2016/v9i28/97780.