Available online at http://www.iorajournal.org/index.php/ijgor/index

International Journal of Global Operations Research

e-ISSN: 2722-1016
p-ISSN: 2723-1739

Vol. 5, No. 2, pp. 119-126, 2024

Application of the Collective Risk Model to the Number of Claims with a Negative Binomial Distribution and the Size of Claims with a Discrete Uniform Distribution

Syavira Syifausufi[1*], Aulianda Anisa Putri[2]

[1,2]*Universitas Padjadjaran, Jatinangor, Indonesia*

**Corresponding author email: syavira20001@mail.unpad.ac.id, aulianda20001@mail.unpad.ac.id*

Abstract

An insurance claim is a form of request from the policy holder to obtain protection against financial losses due to a risk that occurs. Claims that occur every time there is a risk are called individual claims, while the total of individual claims during one insurance period is called aggregate claims. Claims are an important factor in optimizing insurance company expenses, where one of the calculations that insurance companies need to know based on claims is aggregate loss. Aggregate loss is the total loss in a period experienced by policy holders covered by an insurance company. This study aims to determine the average and variance of claims for the number of claims (frequency) with a Negative Binomial distribution and the amount of claims (severity) with a Discreate Uniform distribution in claim payments according to all types of guarantees and the nature of PT injuries. Jasa Raharja (Persero) Purwakarta Representative during the 2018-2020 period. This research uses a collective risk model and the help of Easyfit software to determine the best distribution for the number and size of claims. The results of the research show that from the recapitulation data of claim payments according to all types of coverage and nature of injury in PT. Jasa Raharja (Persero) Purwakarta Representative during the 2018-2020 period, with the number of claims having a Negative Binomial distribution and the amount of claims having a Discrete Uniform distribution, the average aggregate claim occurrence was IDR 2.57818×10^{11} with a variance of IDR 4.22273×10^{21} during the 2018-2020 insurance period.

Keywords: Negative binomial distribution, discrete uniform distribution, aggregate loss claims model, collective risk model

1. Introduction

Humans in living life cannot be separated from risks. Every process in life certainly has uncertainty. If this uncertainty has a bad impact, it is called a risk (Pak, 2014). According to A. Abas Salim, risk is uncertainty which can result in losses. According to Subekti, risk is the obligation to bear losses caused by the occurrence of an event beyond the fault of one of the parties. According to Jorion (1997), there are several types of risk in companies, namely business risk, strategic risk and financial risk. According to Godfrey (1996), there are several sources of risk that need to be known and identified as an initial step in handling risk, namely politics, the environment, planning and marketing.

Insurance is a transfer of risk from the insurance participant to the insurance company as the guarantor under an agreement. In this agreement, insurance participants as insured are required to pay a certain amount of money (premium) as compensation for the risk of loss. Insurance companies must have the ability to anticipate risks during the coverage period and their characteristics. The goal is to estimate the possibility of future losses. With insurance, everyone can reduce the risks that occur (Sukono, 2020).

A claim is a form of demand from the policy holder to the insurance company as the insurance underwriter. According to Carolina (2020), a claim is a form of demand where the insured seeks the right to economic loss protection in accordance with agreed procedures and specified in the policy. Establishing a claims distribution model during the insurance period can be done using two approaches, namely the individual risk model approach and the collective risk model. In the collective risk model, claims that arise every time a risk occurs are called individual claims, while the total of individual claims during one insurance period is called aggregate claims. According to Agustina et al. (2019), aggregate losses are the total losses of policyholders that must be borne by the company in a certain time period. According to Pratiwi et al. (2020), the aggregate loss model is a random variable that has additional benefits in a collection of losses in an insurance policy. Aggregate loss models can be modeled using a collective risk model approach.

The collective risk model denoted by S is a random variable that represents the total amount of loss over many claims and the size of the claim which is distributed i.i.d (independent and identically distributed) which means there is no trend or is taken from the same probability distribution and each sample is an independent event that is not connected to each other (Espinoza, 2021).

According to Rudi et al. (2020), Claim aggregation and claim opportunity distribution can be formed from patterns of the number and size of claims. Discrete probability distributions are widely used to model the number of claims, one example is the Negative Binomial distribution. In contrast, continuous probability distributions are widely used to model claim sizes. In this research, the Discrete Uniform distribution is used to model the size of claims. Based on the description above, this research applies a collective risk model to recapitulation data on claim payments according to all types of coverage and nature of injury in 2018-2020 PT. Jasa Raharja (Persero) Purwakarta Representative. The aim of this research is to determine the average and variance of claims in claim payments according to all types of coverage and nature of PT injuries. Jasa Raharja (Persero) Purwakarta Representative during the 2018-2020 period.

2. Literature Review

Research on collective risk model analysis by forming an aggregation claims model was researched by Saputra,et al. (2018) which examines the analysis of risks borne by insurance companies in an insurance period. This research uses total claims data from companies to form a distribution model of the size and number of claims. The results of this research obtained the average and variance of claims during the insurance period.

Based on several studies that have been carried out previously, this research will apply a collective risk model and claim payment data according to all types of coverage and nature of injury in 2018-2020 PT. Jasa Raharja (Persero) Purwakarta Representative to form a large distribution model and number of claims which was then used to determine the average and variance of claims at insurance companies during the 2018-2020 period.

3. Materials and Methods

3.1. Materials

The data used in this research comes from the article entitled 'Analysis of the Number of Aggregated Claims with Negative Binomial Distribution and the Size of Claims with Discreate Uniform Distribution Using the Convolution Method' published in the journal Journal of Mathematics: Theory and Applications, Volume 1, No. 2, pages 50-56 in 2019. The data used is recapitulation data on claim payments according to all types of guarantees and nature of PT injuries. Jasa Raharja (Persero) Purwakarta Representative during the 2018-2020 period. This data contains the number and size of claims according to all types of coverage and nature of PT injuries. Jasa Raharja (Persero) Purwakarta Representative from January 2018 to December 2020.

3.2. Methods

This research is quantitative research based on calculations and statistical data to determine the relationship with the phenomenon under study. The method used in data collection is the literature and documentation method. The documentation method used is by taking data from claim payments based on all types of coverage and nature of injury from 2018 to 2020. This data is secondary and official data that researchers obtained from the article entitled 'Analysis of the Number of Aggregated Claims with Negative Binomial Distribution and Large Claims with Discreate Uniform Distribution using the Convolution Method' published in the journal Journal of Mathematics: Theory and Applications, Volume 1, No. 2, pages 50-56 in 2019. The literature method used in this research is data and information both through written and electronic documents that can support the writing process.

In carrying out data analysis, the first step taken was to carry out statistical tests and parameter estimates on data on the number of claims and the size of PT insurance claims. Jasa Raharja (Persero) Purwakarta Representative based on all types of coverage and nature of injury in 2018-2020 to determine distribution candidates and their parameters for the number and size of claims. The second step, namely the Kolmogorov-Smirnov test, is carried out to determine a suitable distribution for the data. After obtaining the best distribution for the number and size of claims, the next step is to calculate the average number and size of claims, as well as the variance in the number and size of claims. The final step is to carry out calculations to obtain the average and variance of aggregate claims.

3.2.1. Formula / Equation

Negative Binomial Distribution

The negative binomial distribution is an iterative experiment that runs continuously until a certain number of successes occur (Sudaryono, 2012). The negative binomial distribution is denoted by $b * (x; n, p)$. According to Bruno et al. (2006), If However, because each repeater is independent of each other, it is necessary to multiply by all the probabilities p and failures $q = 1 - p$.

According to Bruno et al. (2006), if a negative binomial experiment has a probability of success p and a probability of failure q, the probability distribution of the random variable

$$p(x) = P(X = x) = \binom{x-1}{n-1} p^n q^{x-n} \tag{1}$$

with:
$p(x)$: probability of a negative binomial distribution with a random variable X and many experimental repetitions up to n times,
x : many tries until you get the nth success,
n : the number of success events that occur,
p : the chance of a successful event occurring,
q : chance of a failure event occurring.

The formula for the mean and variance of the negative binomial distribution is as follows:

$$E(x) = \mu = \frac{n}{p} \tag{2}$$

$$Var(x) = \frac{n(1-p)}{p^2} \tag{3}$$

Discrete Uniform Distribution

The random variable X has a discrete uniform distribution ($X \sim Discrete\ Uniform(a, b)$), indicating that the variable The probability density function (fkp) of the discrete uniform distribution is as follows:

$$f(x) = \frac{1}{b-a+1} \qquad x = a, a+1, \dots, b. \tag{4}$$

with:
$f(x)$: probability if the random variable X has a discrete uniform distribution,
a : first order number,
b : last sequence number.

The formula for the mean and variance of the discrete uniform distribution is as follows:

$$E(x) = \mu = \frac{a+b}{2} \tag{5}$$

$$Var(x) = \frac{(b-a+1)^2 - 1}{12} \tag{6}$$

Aggregate Loss Model

According to Melantika (2023), there are two approaches to modeling aggregate loss S, namely the individual risk model and the collective risk model. The individual risk model emphasizes the losses from each individual contract and expresses the aggregate loss as follows:

$$S_n = X_1 + X_2 + \cdots + X_n \tag{7}$$

withs:
X_i $(i = 1,2, \dots, n)$: the number of losses from a contract of n, at individual risks assumed to be independent,
n : contract amount.

According to Bruno (2006), to obtain aggregate loss, it is done by recording each claim amount and adding up all the claims. Aggregate loss is expressed by a random variable S and the number of claims in one period in a portfolio is expressed by N. The size of each claim can be expressed in random variables $X_1, X_2, \dots$ so that the collective risk model is obtained as follows:

$$S = X_1 + X_2 + \cdots + X_N \qquad N = 0,1,2, \dots \tag{8}$$

According to Klugman et al. (2004), the assumptions that must be considered in aggregate loss for the collective risk model are as follows:

a. Given $N = n$ random variable $X_1, X_2, \dots, X_N$ are random variables that have identical distributions and are independent of each other,
b. Given $N = n$ joint distribution of random variables $X_1, X_2, \dots, X_N$ does not depend on value n,
c. The distribution of the random variable AND does not depend on the values of the random variable $X_1, X_2, \dots, X_N$.

The average occurrence of aggregated claims for the collective risk model is given by:

$$E[S] = E[N]E[X] \tag{9}$$

The variance in the occurrence of aggregate claims in the collective risk model is a conditional variance, namely:

$$Var[S] = E[Var(X|I)] + Var[E(X|I)] \tag{10}$$

If frequency and severity are independent of each other, then the compound variance is:

$$Var[S] = E[N]Var[X] + Var[N]E[X]^2 \tag{11}$$

4. Results and Discussion

In this study, data on total insurance claims based on all types of coverage and the nature of PT injuries were used. Jasa Raharja (Persero) Purwakarta Representative 2018-2020. The number of claims or frequency data shows the number of claim incidents and the severity data or claim size shows the amount of claim payments made by PT. Jasa Rahaja (Persero) Purwakarta Representative as an insurance company. Visualization of frequency and severity data can be seen in Figures 1 and 2.

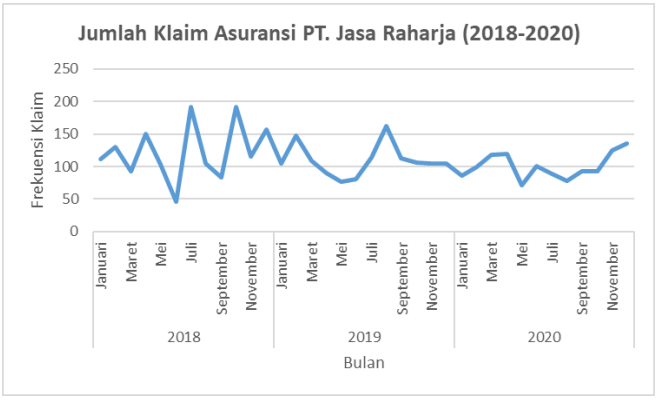

Figure 1: Number of PT Insurance Claims. Raharja Services (2018-2020)

Figure 1 shows a graph of the number of claims or insurance frequency data for PT. Purwakarta Representative Raharja Services experienced an increase and decrease in the number of claims every month during 2018-2020.

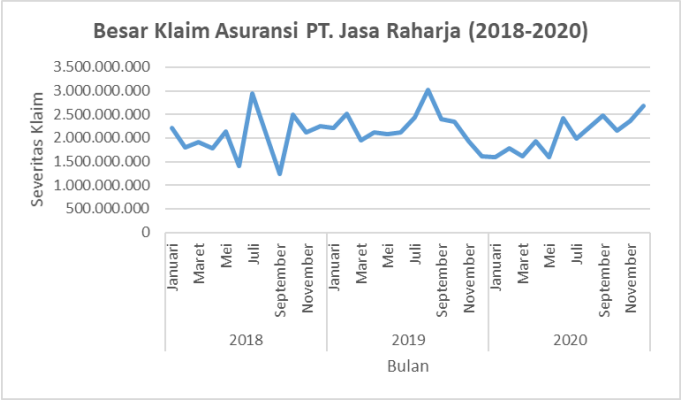

Figure 2: Size of PT Insurance Claims. Raharja Services (2018-2020)

Figure 2 shows a graph of the size of insurance claims or PT severity data. Jasa Raharja (Persero) in 2018-2020 experienced large increases and decreases in claims every month. Then parameter estimation and statistical tests were carried out with the help of Easyfit software to determine the distribution of the number and size of claims according to PT insurance data. Jasa Raharja (Persero) Purwakarta Representative which can be seen in tables 1 and 2.

Table 1: Parameter Estimates for Prospective Distribution of Number of Claims

No	Distribution	Parameter
1	D. Uniform	$a = 57, \quad b = 164$

2	Geometric	$p = 0.00894$
3	Logarithmic	$\theta = 0.99863$
4	Negative Binomial	$n = 14, \quad p = 0.11452$
5	Poisson	$\lambda = 110{,}81$
6	Bernoulli	No fit (data max > 1)
7	Binomial	No fit
8	Hypergeometric	No fit

Based on table 1 which shows the estimated distribution candidate parameters from data on the number of PT insurance claims. Jasa Raharja (Persero) Purwakarta Representative, obtained five distribution candidates who had parameter values from the total of eight distributions obtained. The five candidate distributions are D. Uniform, Geometric, Logarithmic, Negative Binomial, and Poisson. Meanwhile, three other distributions that show No Fit, namely the Bernoulli, Binomial and Hypergeometric distributions, show that there is no match with the data on the number of claims. Then Kolmogorov-Smirnov was carried out to determine whether the data came from a population with a certain distribution or not and obtained table 2 below.

Table 2: Prospective Statistical Values for Distribution of Number of Claims

No	Distribusi	Kolmogorov-Smirnov	
		Statistik	Peringkat
1	D. Uniform	0.15741	2
2	Geometric	0.44854	4
3	Logarithmic	0.69274	5
4	Negative Binomial	0.09945	1
5	Poisson	0.28622	3
6	Bernoulli	No fit (data max>1)	
7	Binomial	No fit	
8	Hypergeometric	No fit	

Table 2 shows the ranking based on the Kolmogorov-Smirnov test, namely Negative Binomial, D. Uniform, Poisson, Geometric, and Logarithmic, while the other three distributions show a mismatch. Then the p-value calculation was carried out to determine the best distribution for the number of claims data and the results obtained were as shown in table 3.

Table 3: P-Value Value of Prospective Distribution of Number of Claims

No	Distribusi	P-Value	Rating
1	Negative Binomial	0.83423	1
2	D. Uniform	0.30177	2
3	Poisson	0.00419	3
4	Geometric	5.1659E-7	4
5	Logarithmic	3.8668E-16	5

Then the Kolmogorov-Smirnov hypothesis test is carried out with,
Hipotesis : H_0: Data comes from a population with a certain distribution
H_1: DData does not come from a population with a certain distribution.
Real level : $\alpha = 5\%$
Test statistics : Kolmogorov-Smirnov test
Test criteria : The $p - value > \alpha = 0.05$ means H_0 is accepted (failed to reject H_0)

Based on the p-value in table 3, the Kolmogorov-Smirnov test results are obtained as in table 4 below:

Table 4: Results of the Kolmogorov-Smirnov Test for Prospective Distribution of Number of Claims

No	Distribution	P-Value	Test results
1	Negative Binomial	0.83423	H_0 accepted
2	D. Uniform	0.30177	H_0 accepted
3	Poisson	0.00419	H_0 rejected
4	Geometric	5.1659E-7	H_0 rejected
5	Logarithmic	3.8668E-16	H_0 rejected

The results of the Kolmogorov-Smirnov test above show that the hypothesis decision for Negative Binomial and D.Uniform is that H_0 is accepted or fails to reject H_0 because the $p-value$ obtained has a $p-value > \alpha$, namely 0.05, which means the data comes from a certain population. Meanwhile, the Poisson, Geometric and Logarithmic distribution candidates have a $p-value < \alpha$ so they reject H_0, which means the data does not come from a population with a certain distribution. Based on the largest p-value and the ranking order of the two candidate distributions previously obtained, the Negative Binomial distribution is the best distribution for modeling the number of PT insurance claims. Jasa Raharja (Persero) Purwakarta Representative.

Next, parameter estimation and statistical tests will be carried out for severity data or claim size data, such as those carried out for frequency data or claim number data. Parameter estimation results for severity data can be seen in table 5.

Table 5: Parameter Estimates for Prospective Claim Size Distribution

No	Distribution	Parameter
1	D. Uniform	a =1.4334E+9 b =2.7845E+9
2	Geometric	p =4.7418E-10
3	Logarithmic	$\theta = 1.0$
4	Negative Binomial	$n = 29,\ p$ =1.3863E-8
5	Poisson	λ =2.1089E+9
6	Bernoulli	No fit (data max > 1)
7	Binomial	No fit
8	Hypergeometric	No fit

Based on table 5 which shows the estimated distribution candidate parameters from PT insurance claims data. Jasa Raharja (Persero) Purwakarta Representative, obtained five distribution candidates who had parameter values from the total of eight distributions obtained. The five candidate distributions, namely D. Uniform, Geometric, Logarithmic, Negative Binomial, and Poisson, while the other three distributions that show No Fit, namely the Bernoulli, Binomial, and Hypergeometric distributions show that they are not suitable for large data claims. Then Kolmogorov-Smirnov was carried out to determine whether the data came from a population with a certain distribution or not and obtained table 6 below.

Table 6: Statistical Value of Prospective Claim Size Distribution

No	Distribution	Kolmogorov-Smirnov	
		Statistics	Rating
1	D. Uniform	0.11204	1
2	Geometric	0.47433	2
3	Logarithmic	N/A	N/A
4	Negative Binomial	N/A	N/A
5	Poisson	0.55556	3
6	Bernoulli	No fit (data max>1)	
7	Binomial	No fit	
8	Hypergeometric	No fit	

Table 6 shows that there are three candidate distributions that have statistical value, while the other five candidate distributions do not have statistical value. Based on the ranking order of the Kolmogorov-Smirnov test, the D. Uniform distribution candidate is ranked first, Geometric is ranked second, and Poisson is ranked third, while Logarithmic and Negative Binomial have N/A results, which means they have no value. Meanwhile, Bernoulli, Binomial and Hypergeometric distributions do not match big climate data or No Fit. Then the p-value calculation was carried out to determine the best distribution for large claims data and the results obtained were as shown in table 7.

Table 7: P-Value Value of Prospective Claim Large Distribution

No	Distribusi	P-Value	Peringkat
1	D. Uniform	0.71502	1
2	Geometric	8.5852E-8	2
3	Poisson	1.5654E-10	3

Then the Kolmogorov-Smirnov hypothesis test is carried out with,
Hypothesis : H_0: Data comes from a population with a certain distribution

H_1: Data does not come from a population with a certain distribution.
Real level : $\alpha = 5\%$
Test statistics : Kolmogorov-Smirnov test
Test criteria : The $p - value\ is\ > \alpha = 0.05$ then H_0 is accepted (failed to reject H_0)

Based on the p-value in table 3, the Kolmogorov-Smirnov test results are obtained as in table 8 below:

Table 8: Kolmogorov-Smirnov Test Results Prospective Claim Size Distribution

No	Distribution	$P - Value$	Test results
1	D. Uniform	0.71502	H_0 accepted
2	Geometric	8.5852E-8	H_0 rejected
3	Poisson	1.5654E-10	H_0 rejected

The results of the Kolmogorov-Smirnov test above show that the hypothesis decision for D.Uniform is that H_0 is accepted or fails to reject H_0 because the p-value obtained has a $p - value\ > \alpha$, namely 0.05, which means the data comes from a certain population. Meanwhile, the Geometric and Poisson distribution candidates have a $p - value < \alpha$ so they reject H_0, which means the data does not come from a population with a certain distribution. Based on the largest $p - value$ and the ranking order obtained previously, the D. Uniform distribution is the best distribution in modeling the size of PT insurance claims. Jasa Raharja (Persero) Purwakarta Representative.

After obtaining the best distribution for frequency data or number of claims, namely the Negative Binomial distribution, and the best distribution for severity data or claim size data, namely the D. Uniform distribution, then the average number of claims, average size of claims, variation can be calculated. number of claims, and variance in claim size.

The number of claims has a Negative Binomial distribution, so the average (mean) is given by equation (2) and the variance is given by equation (3). With the parameters obtained from the parameter estimates above, namely $n = 14$ and $p = 0.11452$, the average and variance of the number of claims can be written as follows:

$$E[N] = \frac{14}{0.11452} = 122.2493888$$

$$Var[N] = \frac{14(1 - 0.11452)}{0.11452^2} = 945.2444006$$

The claim size has a D. Uniform distribution, so the average (mean) is given by equation (5) and the variance is given by equation (6). With the parameters obtained from the parameter estimates above, namely $a = 1.4334E + 9 = 1.4334 \times 10^9$ and $b = 2.7845E + 9 = 2.7845 \times 10^9$, so the average and variance are large the claim can be written as follows:

$$E[X] = \frac{\left(1.4334 \times 10^9\right) + (2.7845 \times 10^9)}{2} = 2108950000$$

$$Var[X] = \frac{(2.7845 \times 10^9 - 1.4334 \times 10^9 + 1)^2 - 1}{12} = 1.52123 \times 10^{17}$$

After obtaining the average and variance of the number and size of claims, the average and variance of aggregate claims can be calculated in that period. In the collective risk model, the average occurrence of aggregated claims is given by equation (9) and the variance of the occurrence of aggregated claims is given by equation (11). So, the results obtained are the average value and variance of the occurrence of claims in claim payments according to all types of guarantees and the nature of PT injuries. Jasa Raharja (Persero) Purwakarta Representatives during the 2018-2020 period are as follows:

$$E[S] = 122.2493888 \times 2108950000 = 2.57818 \times 10^{11}$$

$$Var[S] = 122.2493888 \times 1.52123 \times 10^{17} + 945.2444006 \times (2108950000)^2$$

$$Var[S] = 4.22273 \times 10^{21}$$

Based on this value, the average occurrence of aggregate claims is IDR 2.57818×10^{11} with a variance of IDR 4.22273×10^{21}.

5. Conclusion

Based on the processing above, it was found that the frequency data or number of claims had a Negeative Binomial distribution and the severity data or claim size had a D. Uniform distribution. Based on data calculations, the average aggregate claim occurrence was IDR 2.57818×10^{11} with a variance of IDR 4.22273×10^{21} during the 2018-2020 insurance period.

References

Bruno, M. G., Camerini, E., Manna, A., & Tomassetti, A. (2006). A new method for evaluating the distribution of aggregate claims. *Applied mathematics and computation*, *176*(2), 488-505.

Espinoza, E., Saputri, U., Fadilah, F. H., & Devianto, D. (2021, June). Modeling the count data of public health service visits with overdispersion problem by using negative binomial regression. In *Journal of Physics: Conference Series* (Vol. 1940, No. 1, p. 012021). IOP Publishing.

Godfrey, J. S. (1996). The effect of the Indonesian throughflow on ocean circulation and heat exchange with the atmosphere: A review. *Journal of Geophysical Research: Oceans*, *101*(C5), 12217-12237.

Jorion, P. (1997). In defense of VaR. *Derivatives Strategy*, *2*(4), 20-23.

Kartikasari, M., D. (2017). Premium Pricing of Liability Insurance Using Random Sum Model, 17(1), 46-54.

Pak, R. J. (2014). Estimating loss severity distribution convolution approach. *J. Math. Statist*, *10*(3), 247-254.

Saputra, A., & Rusyaman, E. (2018, March). Risk adjustment model of credit life insurance using a genetic algorithm. In *IOP Conference Series: Materials Science and Engineering* (Vol. 332, No. 1, p. 012007). IOP Publishing.

Sukono, S. S., Mamat, M., & Bon, A. T. (2020, August). Model for determining natural disaster insurance premiums in indonesia using the black scholes method. In *Proceedings of the International Conference on Industrial Engineering and Operations Management, Detroit, Michigan, USA*.