Taksiran Distribusi Frekuensi *Poisson-Tweedie Loss Frequency* dalam Memodelkan Kerugian Agregat Menggunakan *Panjer Recursion* untuk Menentukan *Risk Premium*

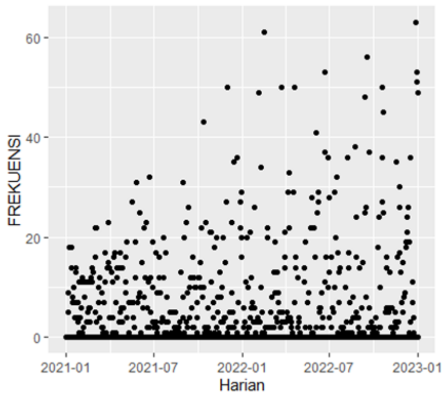
Safa Rahmadila, Galuh Oktavia Siswono, dan Moch Taufik Hakiki
Departemen Aktuaria, Institut Teknologi Sepuluh Nopember (ITS)
e-mail: galuh@aktuaria.its.ac.id

***Abstrak*—Kerugian agregat merupakan total kerugian yang dialami oleh perusahaan asuransi pada periode waktu tertentu. Kerugia ini berasal dari klaim yang diajukan oleh para pemegang polis. Perhitungan kerugian agregat dimodelkan dari frekuensi dan *severity* klaim. Dalam memodelkan frekuensi klaim dari data, dilakukan dengan melihat nilai AIC dengan membandingkan distribusi Poisson, Negatif Binomial, Geometri, dan *Poisson-Tweedie Loss Frequency*. Sedangkan untuk distribusi *severity* dilihat melalui uji Anderson-Darling. Distribusi frekuensi dan *severity* tersebut digunakan dalam memodelkan kerugian agregat. Kemudian, dilakukan perhitungan *risk premium* menggunakan model kerugian agregat menggunakan metode *Panjer Recursion*. Sumber data yang digunakan pada penelitian ini adalah data klaim asuransi kecelakaan PT Jasa Raharja (Persero) KPJR Mojokerto tahun 2021-2022. Data tersebut kemudian dikelompokkan sesuai dengan periode yang digunakan dalam penelitian yaitu harian, mingguan, dan bulanan. Berdasarkan pada analisis periode yang digunakan, periode data yang paling baik dalam pemodelan yaitu data klaim bulanan. Pendekatan distribusi frekuensi menggunakan *Poisson-Tweedie Loss Frequency* dilakukan menggunakan data frekuensi klaim bulanan dan memiliki nilai AIC yang paling kecil jika dibandingkan dengan distribusi Poisson, Negatif Binomial, dan Geometri. Estimasi parameter yang dihasilkan yaitu $\hat{a} = 2,14286$, $\hat{b} = 6,42408 \times 10^3$, dan $\hat{c} = 1,04907$. Sedangkan distribusi yang paling baik dalam memodelkan severity klaim yaitu distribusi Lognormal dengan parameter μ =22,147689 dan $\sigma = 0,23259$. Dari kedua distribusi ini kemudian dibentuk model kerugian agregat yang menghasilkan nilai *risk premium*. Nilai *risk premium* akan bertambah seiring dengan bertambahnya nilai *loading factor*.**

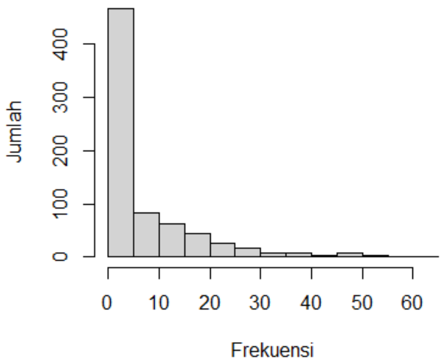
***Kata Kunci*—Kerugian Agregat, *Panjer Recursion*, *Poisson-Tweedie Loss Frequency*, *Risk Premium*.**

I. PENDAHULUAN

PERTUMBUHAN industri asuransi berkembang pesat pada sepuluh tahun terakhir. Asosiasi Asuransi Jiwa Indonesia (AAJI) menyebutkan telah terjadi peningkatan jumlah tertanggung asuransi jiwa sebesar 48,8% jika dibandingkan tahun 2020. Hal ini menunjukkan bahwa kesadaran masyarakat akan pentingnya proteksi diri dari peluang mengalami kerugian di masa depan semakin meningkat. Namun, sebagai lembaga penanggung kerugian yang mungkin dialami oleh tertanggung, tidak menutup kemungkinan perusahaan asuransi juga akan mengalami kerugian. Berdasarkan hasil keputusan Menteri Keuangan nomor 481/KMK.017/1999 mengatakan bahwa klaim dikategorikan sebagai risiko dari suatu perusahaan asuransi. Kerugian atas klaim-klaim yang diajukan oleh pemegang polis pada periode tertentu disebut dengan kerugian agregat.



Gambar 1. *Scatter plot* data frekuensi klaim harian.



Gambar 2. Histogram data frekuensi klaim harian.

Dalam memodelkan kerugian agregat dapat dilakukan dengan menjumlahkan keseluruhan klaim selama periode waktu tertentu. Model kerugian agregat dapat diperoleh melalui beberapa metode seperti Konvolusi, *Panjer Recursion*, *Fast Fourier Transform* [1]. Dalam penelitian ini digunakan metode *Panjer Recursion* dalam membentuk model kerugian agregat. Distribusi Panjer dibagi menjadi Panjer (*a*, *b*, 0) dan Panjer (*a*, *b*, 1). Distribusi kerugian agregat terdiri dari distribusi frekuensi dan *severity* klaim. Metode yang dapat digunakan dalam pemodelan distribusi frekuensi adalah *Poisson-Tweedie Loss Frequency*. Distribusi *Poisson-Tweedie* memiliki tiga parameter (*a*, *b*, *c*). Distribusi *Poisson-Tweedie* diperkenalkan sebagai perluasan dari distribusi Poisson, Binomial Negatif, dan Pollya-Aeppli [2]. Sedangkan untuk distribusi *severity* klaim dapat dimodelkan dengan distribusi kontinu dengan support selang $[0, \infty)$. Dari model kerugian agregat yang terbentuk, maka dapat dihitung besar *risk premium*. *Risk premium*, yaitu premi

Tabel 1.
Nilai (*a, b*) untuk kelas (*a, b, 0*)

Distribusi	a	b	p_0
Poisson	0	λ	$e^{-\lambda}$
Binomial	$\frac{-p}{1-p}$	$(n+1)\frac{-p}{1-p}$	$(1-p)^n$
Binomial Negatif	$\frac{\beta}{1+\beta}$	$(r-1)\frac{\beta}{1+\beta}$	$(1+\beta)^{-r}$
Geometri	$\frac{\beta}{1+\beta}$	0	$(1+\beta)^{-1}$

Tabel 2.
Nilai (*a, b*) untuk kelas (*a, b, 1*)

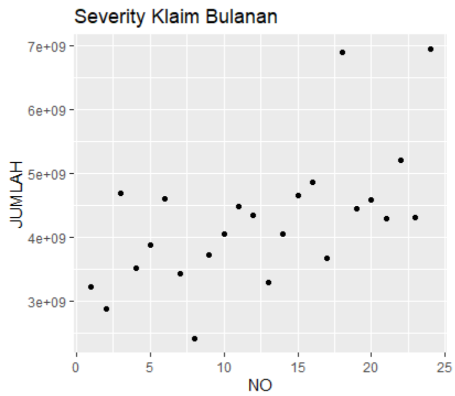
Distribusi	a	b	p_0	Ukuran parameter
Poisson	0	λ	$e^{-\lambda}$	$\lambda > 0$
ZT Poisson	0	λ	0	$\lambda > 0$
ZM Poisson	0	λ	-	$\lambda > 0$
Binomial	$\frac{-p}{1-p}$	$(n+1)\frac{-p}{1-p}$	$(1-p)^n$	$0 < q < 1$
ZT Binomial	$\frac{-p}{1-p}$	$(n+1)\frac{-p}{1-p}$	0	$0 < q < 1$
ZM Binomial	$\frac{-p}{1-p}$	$(n+1)\frac{-p}{1-p}$	-	$0 < q < 1$
Binomial Negatif	$\frac{\beta}{1+\beta}$	$(r-1)\frac{\beta}{1+\beta}$	$(1+\beta)^{-r}$	$r > 0, \beta > 0$
ETNB	$\frac{\beta}{1+\beta}$	$(r-1)\frac{\beta}{1+\beta}$	0	$r > 0,^b \beta > 0$
ZM ETNB	$\frac{\beta}{1+\beta}$	$(r-1)\frac{\beta}{1+\beta}$	-	$r > 0,^b \beta > 0$
Geometri	$\frac{\beta}{1+\beta}$	0	$(1+\beta)^{-1}$	$\beta > 0$
ZT Geometri	$\frac{\beta}{1+\beta}$	0	0	$\beta > 0$

Tabel 3.
Data klaim asuransi kecelakaan

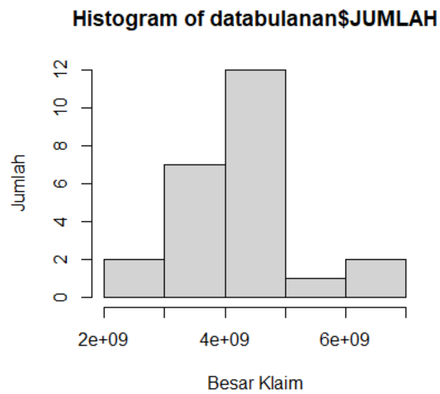
No	Tanggal	Besar Klaim (Rp)	Frekuensi Klaim
1	1 Januari 2021	0	0
2	2 Januari 2021	0	0
3	3 Januari 2021	0	0
4	4 Januari 2021	4.845.483	1
⋮	⋮	⋮	⋮
5.176	30 Desember 2022	2.942.314	1
5.177	31 Desember 2022	0	0

yang dibayarkan setelah ditambahkan faktor loadings [3]. Faktor *loadings* yang dimaksud yaitu biaya yang dikeluarkan untuk membuat satu polis asuransi seperti biaya administrasi.

Pemodelan kerugian agregat dalam menghitung premi murni pernah dilakukan oleh Manurung dan Mannohas (2016). Dalam tulisannya Tohap dan Mans menggunakan metode *Fast Fourier Transform* dalam memodelkan distribusi kerugian agregat yang digunakan untuk menghitung premi murni. Chen (2016) juga melakukan penelitian mengenai penggunaan *Poisson-Tweedie Loss Frequency* dalam membentuk model kerugian agregat. Kemudian, penentuan distribusi frekuensi kerugian agregat menggunakan metode Rekursif Panjer dilakukan oleh Utami, Mutaqin, dan Wachidah (2017). Dari tulisan ini, Dianti, dkk. (2017), menyimpulkan bahwa metode Rekursif Panjer metode yang bersifat numerik serta dapat menjadi solusi apabila pemodelan kerugian agregat menggunakan konvolusi tidak dapat ditemukan. Dari beberapa penelitian yang telah



Gambar 3. *Scatter plot* data *severity* klaim bulanan.



Gambar 4. Histogram data *severity* klaim bulanan.

dilakukan, penulis tertarik untuk menggunakan distribusi *Poisson-Tweedie* dalam memodelkan distribusi frekuensi kerugian agregat. Dari taksiran frekuensi ini nantinya akan dibentuk model kerugian agregat menggunakan *Panjer Recursion* yang digunakan pada perhitungan *risk premium*.

II. TINJAUAN PUSTAKA

A. Asuransi

Menurut Pasal 1 UU No 2/1992, asuransi atau pertanggungan adalah perjanjian antara dua pihak atau lebih, yang mana penanggung mengikatkan diri kepada tertanggung dengan menerima premi asuransi, untuk memberikan penggantian kepada tertanggung karena kerugian, kerusakan atau kehilangan keuntungan yang diharapkan, atau tanggung jawab hukum kepada pihak ketiga yang mungkin akan diderita tertanggung yang timbul dari suatu peristiwa yang tidak pasti atau untuk memberikan pembayaran yang didasarkan atas meninggal atau hidupnya. Terdapat beberapa unsur asuransi seperti klaim, premi, serta polis asuransi.

B. Model Distribusi Frekuensi

Pemodelan distribusi frekuensi atau banyaknya klaim yang terjadi dapat dilakukan melalui distribusi counting, merujuk pada sifat frekuensi yang tidak pernah negatif. Untuk memodelkan distribusi frekuensi, dipilih model distribusi *Binomial*, *Poisson*, *Binomial Negatif*, dan *Geometri* karena memiliki *support* tak negative.

1) Distribusi Binomial

Fungsi massa peluang (*probability mass function* atau pmf) yang dibentuk dari distribusi binomial adalah sebagai berikut:

$$p_k = \mathbb{P}(N = k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \tag{1}$$

Tabel 7.
Data klaim harian

No	Tanggal	Jumlah Besar Klaim (Rp)	Jumlah Frekuensi Klaim
1	1-1-2021	0	0
2	2-1-2021	0	0
3	3-1-2021	0	0
4	4-1-2021	174.845.483	5
⋮	⋮	⋮	⋮
729	30-12-2022	573.246.045	49
730	31-12-2022	0	0

Tabel 8.
Data klaim mingguan

No	Minggu ke-	Jumlah Besar Klaim (Rp)	Jumlah Frekuensi Klaim
1	1	313.572.803	14
2	2	1.486.188.381	65
⋮	⋮	⋮	⋮
103	103	493.965.092	27
104	104	3.115.772.463	216

Tabel 9.
Data klaim bulanan

No	Bulan ke-	Jumlah Besar Klaim (Rp)	Jumlah Frekuensi Klaim
1	1	3.228.500.941	157
2	2	2.884.226.872	146
⋮	⋮	⋮	⋮
23	23	4.309.470.073	207
24	24	6.944.402.578	427

dengan $k = 0, 1, 2, \cdots, n$.

Mean dan variansi dari *N* yang berdistribusi binomial adalah:

$$E(N) = np \tag{2}$$

$$Var(N) = np(1-p) \tag{3}$$

2) Distribusi Poison

Distribusi Poisson memiliki pmf sebagai berikut:

$$p_k = \frac{\lambda^k e^{-\lambda}}{k!} \tag{4}$$

dengan $k = 0, 1, 2, 3, ...$

Mean dan Variansi dari distribusi Poisson dapat dihitung berdasarkan fungsi pembangkit peluangnya yaitu:

$$E(N) = \lambda \tag{5}$$

$$Var(N) = \lambda \tag{6}$$

3) Distribusi Binomial Negatif

Pmf dari distribusi binomial negatif dengan parameter (r, β) adalah:

$$p_k = \mathbb{P}(N = k) = \binom{k+r-1}{k} \left(\frac{1}{1+\beta}\right)^r \left(\frac{\beta}{1+\beta}\right)^k \tag{7}$$

dengan $k = 0, 1, 2, 3, \ldots;\ r > 0;\ \beta > 0$.

Mean dan variansi dari distribusi ini adalah:

$$E(N) = r\beta \tag{8}$$

$$Var(N) = r\beta(\beta + 1) \tag{9}$$

Terdapat kasus khusus pada distribusi binomial negatif, yaitu apabila nilai $r = 1$ maka disebut dengan distribusi Geometri dengan parameter (β).

Tabel 4.
Statistika deskriptif data frekuensi klaim

	Data Harian	Data Mingguan	Data Bulanan
Jumlah	730	104	24
Minimal	0	1	79
Kuartil Bawah	0	27,75	165,2
Median	2	44,50	211,5
Mean	7,089	49,48	215,6
Kuartil Atas	10,750	61,25	244,2
Maksimal	63	216	427
Varians	113,585	1261,65	5483,6

Tabel 5.
Koefisien autokorelasi

	Data Mingguan	Data Bulanan
Durbin-Watson	1,9079	1,6218
Breusch-Godrey	0,5189	0,0001

Tabel 6.
Nilai *confidence interval*

Data Bulanan	
Batas bawah 95% *CI* $\hat{\boldsymbol{a}}$	2,1428
Batas atas 95% *CI* $\hat{\boldsymbol{a}}$	2,1429

4) Kelas (a, b, 0)

Nilai parameter *a* dan *b* untuk masing-masing distribusi ditunjukkan pada Tabel 1.

5) Kelas (a, b, 1)

Nilai parameter *a* dan *b* untuk masing-masing distribusi terdapat pada Tabel 2.

C. Model Distribusi Severity

Pada pemodelan distribusi *severity* (besar klaim) digunakan distibusi kontinu dengan *support* selang $[0, \infty)$, karena nilai klaim yang besar jarang muncul maka distribusi yang digunakan haruslah *right tail* [4].

1) Distribusi Gamma

Jika diberikan *X* peubah acak yang berdistribusi *gamma*, maka fungsi kepadatan peluang dari distribusi ini ditunjukkan pada persamaan berikut:

$$f(x) = \frac{\left(\frac{x}{\theta}\right)^\alpha e^{-\frac{x}{\theta}}}{x\Gamma(\alpha)}, \qquad x \geq 0 \tag{10}$$

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx, \qquad \alpha > 0 \tag{11}$$

Rataan dan variansi dari distribusi gamma adalah sebagai berikut:

$$E(X) = \alpha\theta \tag{12}$$

$$Var(X) = \alpha\theta^2 \tag{13}$$

Kasus khusus pada distribusi ini terjadi apabila nilai $\alpha = 1$, maka distribusi gamma dikenal dengan distribusi eksponensial.

2) Distribusi Lognormal

Distribusi ini memiliki dua parameter yaitu μ dan σ. Fungsi kepadatan peluang dari distribusi ini adalah sebagai berikut:

$$f(y) = \frac{1}{y\,\sigma\sqrt{2\pi}} exp\left(\frac{-z^2}{2}\right) \tag{14}$$

dengan $z = \frac{\ln y - \mu}{\sigma}$.

Tabel 14.
Nilai AIC

Distribusi	AIC
Poisson-Tweedie Loss	274,50
Poisson	726,65
Geometri	308,04
Negative Binomial	274,52

Tabel 15.
Statistika deskriptif data *severity* klaim bulanan

Data Bulanan	
Jumlah	24
Minimal	Rp2.418.283.403
Kuartil Bawah	Rp3.642.180.933
Median	Rp4.303.334.848
Mean	Rp4.270.802.723
Kuartil Atas	Rp4.618.836.147
Maksimal	Rp6.944.402.578
Skewness	0,9960792
Varians	$1.106.884 \times 10^{18}$

Tabel 16.
Nilai estimasi parameter distribusi *severity*

	Gamma	Lognormal	Weibull	Pareto
Shape	18,42737	22,147689	4,10857	1,84718
Scale	$2,31764 \times 10^8$	0,23259	$4,66668 \times 10^9$	$2,41828 \times 10^9$

Rataan dan variansi dari distribusi ini adalah:

$$E[X] = e^{\left(\mu+\frac{1}{2}\sigma^2\right)} \quad (15)$$

$$Var[X] = e^{2\mu+\sigma^2}\left(e^{\sigma^2}-1\right) \quad (16)$$

3) Distribusi Weibul

Misalkan *X* adalah peubah acak yang berdistribusi *Weibull*, maka fungsi kepadatan peluang diberikan oleh:

$$f(x) = \frac{\tau\left(\frac{x}{\theta}\right)^{\tau} e^{-\left(\frac{x}{\theta}\right)^{\tau}}}{x}, \qquad x > 0 \quad (17)$$

Rataan dan variansi dari distribusi *Weibull* ditentukan oleh persamaan berikut:

$$E(X) = \theta\, \Gamma\left(1+\frac{1}{\tau}\right) \quad (18)$$

$$Var(X) = \theta^2 \left\{\Gamma\left(1+\frac{2}{\tau}\right) - \left(\Gamma\left(1+\frac{1}{\tau}\right)\right)^2\right\} \quad (19)$$

4) Distribusi Pareto

Misalkan *X* merupakan peubah acak dengan distribusi *Pareto*, maka fungsi kepadatan peluang adalah sebagai berikut:

$$f(x) = \frac{\alpha\, \theta^{\alpha}}{(x+\theta)^{\alpha+1}} \quad (20)$$

D. Metode Maximum Likelihood Estimation

Dalam menentukan parameter distribusi menggunakan metode *MLE*, data terdiri dari *n* percobaan yang berasal dari *n* peubah acak $X_1, X_2, \cdots, X_N$. Nilai $X_1, X_2, \cdots, X_N$ tidak harus berasal dari satu distribusi, namun harus dipastikan bahwa nilai tersebut memiliki parameter yang sama.

$$L(\theta) = \prod_{j=1}^{n} f_X\left(x_j \middle| \theta\right) \quad (21)$$

Tabel 10.
Statistik uji *anderson darling*

	Gamma	Lognormal	Weibull	Pareto
Anderson-Darling	0,505799	0,441750	1,011492	0,389424
p-value	0,182499	0,265121	0,009433	0,355205

Tabel 11.
Nilai AIC distribusi *serverity*

Distribusi	AIC
Gamma	1065,701
Lognormal	1065,191
Pareto	1085,633

Tabel 12.
Ekspektasi dan variansi model kerugian agregat

Severity klaim		Frekuensi klaim		$E(S)$	$Var(S)$
$E(X)$	$Var(X)$	$E(N)$	$Var(N)$		
$4,26939 \times 10^9$	$1,82277 \times 10^{19}$	215,625	5483,636	$9,20588 \times 10^{11}$	$1,03885 \times 10^{23}$

Tabel 13.
Nilai *risk premium*

m	*Risk Premium*		
	Expectation Principle	*Standar Deviation Principle*	*Variance Principle*
1	$1,84117 \times 10^{12}$	$1,24289 \times 10^{12}$	$1,03885 \times 10^{23}$
1,5	$2,30147 \times 10^{12}$	$1,40405 \times 10^{12}$	$1,55827 \times 10^{23}$
2	$2,76176 \times 10^{12}$	$1,56521 \times 10^{12}$	$2,07769 \times 10^{23}$

Untuk menentukkan parameter menggunakan *MLE*, yang harus dilakukan adalah mencari turunan pertama dari $l(\theta) = \ln L(\theta)$, yaitu $\frac{\partial l(\theta)}{\partial \theta} = 0$, sehingga didapatkan nilai θ yang memaksimumkan fungsi likelihood.

E. Uji Kecocokan Model

1) Uji Anderson-Daling

Asumsi yang digunakan pada uji Anderson-Darling yaitu tidak ada parameter yang akan diperkirakan dalam distribusi yang diuji. Statistik ujinya ditunjukkan pada persamaan di bawah ini:

$$A = -n - \frac{1}{n}\sum_{i=1}^{n}[(2i-1)\cdot ln(Ei) + (2(n-i)+1)\cdot ln(1-Ei)] \quad (22)$$

2) Confidence Interval

Selang kepercayaan (*confidence interval*) merupakan metode yang digunakan untuk mengukur keakuratan parameter populasi berdasarkan sampel menggunakan metode statistik tertentu dalam sebuah rentang tertentu.

$$\hat{a} \pm C_{\alpha} \times \frac{SE(\hat{a})}{\sqrt{T}} \quad (23)$$

3) Uji Autokorelasi

Autokorelasi merupakan korelasi yang terjadi antar residual pada satu pengamatan dengan pengamatan lainnya pada model regresi [5]. Metode yang paling umum digunakan yaitu Durbin-Watson (DW) dengan statistik uji sebagai berikut:

$$d = \frac{\sum_{t=2}^{n}(\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^{n}\varepsilon_t{}^2} \quad (24)$$

Selain uji Durbin-Watson, pengujian autokorelasi juga dapat dilakukan dengan metode Breusch-Godfrey.

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \cdots + \rho_p \varepsilon_{t-1} + v_t \quad (25)$$

4) Akaike Information Criterion

Pemilihan model terbaik ditentukan berdasarkan nilai AIC minimum menggunakan persamaan sebagai berikut:

$$AIC = -2l + 2k \quad (26)$$

F. Pemilihan Model

Pemilihan model dilakukan setelah uji hipotesis, yang menggunakan metode *score-based approach*. Metode ini dilakukan dengan memberikan penilaian untuk model-model yang dirasa paling cocok berdasarkan pada hasil pengujian hipotesis yang telah dilakukan sebelumnya. Kriteria model didasarkan pada hal-hal berikut ini:

1. Nilai statistik uji terendah pada uji Anderson-Daring.
2. Nilai koefisien korelasi terendah pada uji Autokorelasi.
3. Nilai AIC terendah.

G. Model Kerugian Agregat

Aggregate loss atau kerugian agregat merupakan total kerugian yang dialami oleh perusahaan asuransi dalam suatu periode waktu tertentu, kerugian tersebut bersumber dari klaim yang diajukan oleh pemegang polis [6]. Distribusi kerugian agregat terdiri dari distribusi frekuensi dan distribusi *severity* klaim, sehingga distribusi kerugian agregat sering disebut dengan distribusi *compound* (*compound distribution*). Kerugian agregat dapat dinyatakan dengan *S*, dari *N* jumlah klaim, dan dari pembayaran polis asuransi $(X_1, X_2, \cdots, X_N)$ [7].

$$S = X_1 + X_2 + \cdots + X_N \quad (27)$$

Model di atas disebut dengan *collective risk models*, dimana X_j peubah acak yang identik dan saling bebas (*independent and identically distributed* atau *i.i.d*) dengan asumsi-asumsi sebagai berikut:

1. Diberikan $N = n$, peubah acak $X_1, X_2, \cdots, X_N$ haruslah *i.i.d.*
2. Diberikan $N = n$, peubah acak $X_1, X_2, \cdots, X_N$ tidak bergantung pada *n*.
3. Distribusi dari *N* tidak bergantung pada nilai $X_1, X_2, \cdots, X_N$.

H. Kerugian Agregat dengan Model Compound

Peluang tepat terjadinya total klaim sebesar *k* adalah sebagai berikut:

$$P(S = k) = \sum_{n=0}^{\infty} \mathbb{P}(X_1 + X_2 + \cdots + X_N = k)\,\mathbb{P}(N = n) \quad (28)$$

Apabila dimisalkan $g_n = \mathbb{P}(S = k), p_n = \mathbb{P}(N = n)$, dan $f_n = \mathbb{P}(X = n)$, maka persamaan di atas dapat disederhanakan menjadi:

$$g_n = \sum_{n=0}^{\infty} p_n f_k^{*n} \quad (29)$$

Jika *X* merupakan peubah acak diskrit dengan peluang bernilai positif pada 0, 1, 2, 3, …, maka:

$$F_X^{*k}(x) = \sum_0^x F_X^{*(k-1)}(x - y) f_X(y) dy \quad (30)$$

dengan $x = 0, 1, 2, \ldots, k = 2, 3, \ldots.$

pdf-nya ditulis pada persamaan berikut:

$$f_X^{*k}(x) = \sum_0^x f_X^{*(k-1)}(x - y) f_X(y) dy \quad (31)$$

dengan $x = 0, 1, 2, \ldots, k = 2, 3, \ldots.$

Fungsi kepadatan peluang dari kerugian agregat ditunjukkan pada persamaan berikut:

$$f_s(x) = \sum_{n=0}^{\infty} p_n f_X^{*k}(x) \quad (32)$$

Rataan dan variansi dari *S* ditunjukkan pada persamaan berikut:

$$E(S) = E(N) \times E(X) \quad (33)$$

$$Var(S) = E(N)Var(X) + Var(N)[E(X)]^2 \quad (34)$$

I. Panjer Recursion

Suatu metode rekursi dikenalkan oleh Panjer, sehingga dikenal sebagai *Panjer Recursion*. *Panjer recursion* hanya bisa digunakan pada nilai besar klaim yang diskrit. Jadi, untuk mengaplikasikan metode ini, *severity* pada klaim harus didiskritkan terlebih dahulu. Distribusi *compound* kelas Panjer (*a*, *b*, 0) ditunjukkan pada persamaan berikut:

$$f_s(x) = \frac{1}{1 - a\, f_X(0)} \sum_{y=1}^{x \wedge m} \left(a + \frac{b\, y}{x}\right) f_X(y)\, f_s(x - y) \quad (35)$$

Sedangkan, untuk distribusi *compound* dari kelas Panjer (*a*, *b*, 1) mengikuti persamaan berikut:

$$f_S(x) = \frac{[p_1 - (a+b)p_0] f_X(x) + \sum_{y=1}^{x \wedge m}\left(a + \frac{b\, y}{x}\right) f_X(y)\, f_S(x-y)}{1 - a\, f_X(0)} \quad (36)$$

J. Diskritisasi

Dalam pengaplikasian metode *Panjer Recursion*, besarnya klaim (*severity*) haruslah berdistribusi diskrit, sehingga perlu dilakukan pendiskritan distribusi besar klaim melalui metode yang disebut dengan *method of rounding* [8]. Dari konstruksi ini, akan dihasilkan ruang sampel bilangan bulat non negatif dengan nilai di setiap ruang sampelnya mewakili besar klaim tertentu yang apabila dikalikan dengan suatu bilangan pengali yang sama (*h*) akan menghasilkan nilai besar klaim yang diwakilinya.

$$f(0) = \mathbb{P}\left(X < \frac{h}{2}\right) = F_X\left(\frac{h}{2}\right) \quad (37)$$

$$f_j = F_X\left(\left(j + \frac{1}{2}\right)h\right) - F_X\left(\left(j - \frac{1}{2}\right)h\right) \quad (38)$$

K. Risk Premium

Dalam penelitian kali ini, akan dilakukan *perhitungan risk premium* berdasarkan model kerugian agregat yang terbentuk. *Risk premium* adalah biaya yang dibebankan oleh perusahaan asuransi kepada pemegang polis yang ditambahkan faktor *loading*. Faktor *loading* (*m*) adalah biaya yang dikeluarkan untuk membuat satu polis asuransi seperti biaya administrasi atau biaya operasional lain. Dalam tulisan ini digunakan metode *risk measure* dalam *premium principle*. Berikut merupakan tiga tipe dari *risk measure* dalam menetapkan premi.

1. *Expectation principle*

$$\Pi(X) = (1 + m)\mathbb{E}[X]\ ,\ m > 0 \quad (39)$$

2. *Standar deviation principle*

$$\Pi(X) = \mathbb{E}[X] + m\sqrt{Var[X]} \ \ , \ \ m > 0 \qquad (40)$$

3. *Variance principle*

$$\Pi(X) = \mathbb{E}[X] + m(Var[X]) \ \ , \ \ m > 0 \qquad (41)$$

III. METODOLOGI PENELITIAN

A. Sumber Data

Penelitian ini menggunakan sekunder yang diperoleh dari PT Asuransi Jasa Raharja (Persero) KPJR Mojokerto. Data yang digunakan berupa data klaim asuransi kendaraan bermotor dengan tahun data 2021 hingga 2022.

B. Langkah Analitis

Langkah-langkah analisis pada penelitian ini diantaranya yaitu sebagai berikut:

1. Melakukan *preprocessing* data klaim asuransi kecelakaan untuk variabel *severity* dan frekuensi klaim per periode (harian, mingguan, bulanan).
2. Menentukan periode terbaik dalam memodelkan data klaim asuransi.
3. Menentukan distribusi terbaik dalam memodelkan data frekuensi klaim dengan membandingkan nilai AIC masing-masing distribusi.
4. Menentukan distribusi terbaik dalam memodelkan data *severity* klaim melalui uji Anderson-Darling.
5. Membentuk model kerugian agregat berdasarkan parameter yang telah didapatkan dari distribusi frekuensi dan *severity*.
6. Menghitung *risk premium* berdasarkan tiga *premium principle*.
7. Menghasilkan kesimpulan.

IV. HASIL DAN PEMBAHASAN

A. Menentukan Periode Distribusi Klaim

Penelitian ini menggunakan data frekuensi dan *severity* klaim seperti yang ditunjukkan pada Tabel 3.

Data klaim asuransi kecelakaan secara keseluruhan kemudian dikelompokkan sesuai dengan periode yang digunakan dalam penelitian yaitu klaim harian yang ditunjukkan pada Tabel 4, klaim mingguan yang ditunjukkan pada Tabel 5, dan klaim bulanan yang ditunjukkan pada Tabel 6.

Maka langkah selanjutnya yaitu mencari statistika deskriptif seperti yang ditampilkan pada Tabel 7 serta membentuk *scatter plot* dan histogram menggunakan data frekuensi klaim seperti yang ditampilkan pada Gambar 1 dan Gambar 2.

Pada *scatter plot* harian, dapat dilihat bahwa frekuensi klaim berjumlah 0 banyak terjadi, yang berarti bahwa data frekuensi klaim harian memiliki *zero-inflated* yang tinggi. Nilai 0 pada data frekuensi klaim menunjukkan bahwa tidak ada klaim yang terjadi di waktu tersebut, data dengan jumlah 0 yang lebih banyak dinilai kurang mampu memodelkan kerugian agregat [2]. Sehingga, dapat disimpulkan bahwa data frekuensi klaim periode harian kurang baik dalam memodelkan data. Sedangkan untuk data frekuensi klaim mingguan dan bulanan akan dilakukan pengujian autokorelasi untuk menentukan periode terbaik dalam memodelkan data. Pada penelitian ini, pengujian autokorelasi dilakukan dengan metode Durbin-Watson serta Breusch-Godfrey yang hasilnya ditampilkan pada Tabel 8.

B. Analisis Distribusi Frekuensi

Pemodelan distribusi frekuensi klaim pada penelitian ini akan menggunakan *Poisson-Tweedie Frequency*. Berikut merupakan *mean*, varians, serta nilai *Value at Risk* dari data klaim bulanan. Data frekuensi klaim bulanan memiliki *mean* sebesar 215,625 dengan varians sebesar 5483,636. Nilai *mean* yang lebih kecil dibandingkan dengan varians menunjukkan adanya *overdispersion* pada data. *Overdispersion* merujuk pada keadaan dimana varians data lebih besar daripada yang diharapkan oleh model yang diasumsikan. Salah satu distribusi yang dapat memodelkan data dengan *overdispersion* adalah distribusi *Poisson-Tweedie.* Nilai $VaR_{95\%}$sebesar 146 menunjukkan bahwa dengan tingkat kepercayaan 95%, jumlah klaim yang terjadi tiap bulan kurang dari 146. Kemudian, dengan metode *Maximum Likelihood Estimation* maka dapat diperkirakan nilai dari parameter $\hat{a}$ untuk data frekuensi klaim sebesar 2,142857 dan standar error sebesar 0,000172 Dari nilai parameter $\hat{a}$ tersebut, dapat ditentukan nilai *confidence interval* dengan persamaan 23.

$$CI = 2{,}142857 \pm C_{rit\alpha/_2} \times \frac{0.000172}{\sqrt{24}}$$

T merupakan jumlah periode yang digunakan yaitu 24, $C_{rit\alpha/_2}$merupakan *95% critical value* berdasarkan asumsi distribusi dari estimasi parameter, sesuai dengan pembahasan pada subbab 2.6.2, karena nilai $T \leq 50$ maka menggunakan distribusi *Student-t.* Karena menggunakan tabel distribusi *Student-t* maka nilai $df = 24$ dengan $significance\ level =$ $0{,}05.$ Maka nilai *critical value* $t_{1-\alpha/_2} = 2{,}069$. Nilai *CI* ditunjukkan pada Tabel 9.

Dalam keluarga *Poisson-Tweedie*, parameter $\hat{a}$ untuk setiap distribusi memiliki nilai yang berbeda. Untuk distribusi Poisson sebesar $\hat{a} = 1$, distribusi negatif binomial $\hat{a} = 0$, dan distribusi Poisson-Inverse Gaussian memiliki nilai $\hat{a} = 0{,}5$. Karena nilai 95% *confidence interval* pada Tabel 10 bernilai positif, maka ketiga distribusi tersebut tidak cocok dalam memodelkan data.

Dari Tabel 10 dapat diketahui bahwa nilai AIC untuk masing-masing distribusi. Karena distribusi *Poisson-Tweedie* memiliki AIC terkecil, maka dapat disimpulkan bahwa distribusi yang paling baik dalam memodelkan data frekuensi klaim adalah distribusi *Poisson-Tweedie.*

Kemudian, dari nilai parameter $\hat{a}$ yang ada dapat ditentukan nilai parameter $\hat{b}$ dan juga $\hat{c}$ dengan persamaan ekspektasi dan variansi sebesar $\mathbf{6{,}42408 \times 10^3}$ dan 1,04907.

C. Analisis Distribusi Severity

Statistika deskriptif data *severity* klaim bulanan ditunjukkan pada Tabel 11.

Kemudian, dari data *severity* klaim bulanan akan dibentuk *scatter plot* seperti pada Gambar 3 serta histogram seperti pada Gambar 4 untuk menentukan distribusi yang sesuai memodelkan data.

Dari Tabel 11, diketahui bahwa *skewness* untuk data *severity* klaim memiliki nilai yang positif sebesar 0,9960792.

Sehingga, data *severity* klaim dapat dimodelkan menggunakan distribusi yang *skewed to the right*. Beberapa distribusi *skewed to the right* yaitu distribusi Gamma, Lognormal, Weibull, serta Pareto. Hasil estimasi parameter masing-masing distribusi ditunjukkan pada Tabel 12.

Setelah mendapatkan nilai taksiran parameter, langkah selanjutnya yaitu melakukan uji Anderson-Darling. Nilai statistik uji Anderson-Darling dapat dilihat pada Tabel 13.

Karena pada distribusi Weibull nilai p-value $< 0{,}05$, maka tolak H_0 yang berarti bahwa data tidak berdistribusi Weibull. Kemudian, dilakukan pengujian kembali untuk distribusi gamma, lognormal, dan pareto menggunakan AIC. Nilai AIC masing-masing distribusi ditunjukkan pada Tabel 14.

Karena nilai AIC terkecil adalah distribusi lognormal, maka distribusi yang paling sesuai dalam memodelkan data *severity* klaim adalah distribusi Lognormal. Distribusi Lognormal memiliki dua parameter yaitu yaitu μ dan σ. Berdasarkan Tabel 12 diketahui bahwa parameter μ atau meanlog dari distribusi ini sebesar 22,147689, sedangkan parameter σ atau standar deviation log sebesar 0,23259. Fungsi kepadatan peluang data *severity* klaim yang berdistribusi lognormal dapat dituliskan sebagai berikut:

$$f(y) = \frac{1}{y(0{,}23259)\sqrt{2(22{,}147689)}} exp\left(\frac{-z^2}{2}\right)$$

dimana

$$z = \frac{\ln y - (22{,}147689)}{(0{,}23259)}$$

Ekspektasi dan variansi dari distribusi Lognormal adalah sebagai berikut:

$$E[X] = 4{,}26939 \times 10^9$$

$$Var[X] = 1{,}82277 \times 10^{19}$$

D. Diskritisasi

Dalam menentukan model kerugian agregat menggunakan metode rekursif perlu dilakukan proses diskritisasi terhadap *severity* klaim terlebih dahulu. Metode yang akan digunakan yaitu *method of rounding*. Misalkan bilangan pengali yang digunakan yaitu $h = 1.000.000.000$, maka nilai ruang sampel peubah acak Y mewakili besar klaim sebesar 1.000.000.000 atau 10^9 dikalikan dengan nilai tersebut. Untuk menghitung fungsi peluang dari peubah acak Y, terlebih dahulu dihitung peluang Y di titik nol.

$$f_Y(0) = \mathbb{P}\left(X < \frac{1.000.000.000}{2}\right) = F_X(500.000.000)$$

$$\begin{aligned} f_Y(j) &= \mathbb{P}(1.000.000.000j - 500.000.000 \le X \\ &\le 1.000.000.000j + 500.000.000) \\ &= F_X(1.000.000.000j - 500.000.000) \\ &- F_X(1.000.000.000j + 500.000.000) \end{aligned}$$

Kemudian, dicari nilai peluang distribusi aritmatika *severity* klaim di titik nol yang diperoleh dari perhitungan di bawah ini.

$$\begin{aligned} f_Y(0) = F_X(500.000.000) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{ln\,500.000.000-(22{,}147689)}{(0{,}23259)}\right)^2} \\ &= 3{,}9987897708 \times 10^{-19} \end{aligned}$$

E. Menentukan Model KErugian Agregat

Kerugian agregat merupakan total kerugian yang harus ditanggung oleh perusahaan asuransi. Dalam hal ini, model kerugian agregat adalah total biaya klaim kecelakaan yang ditanggung perusahaan atas klaim peserta untuk *underwriting year* 2021 hingga 2022. Model kerugian agregat menggunakan *Panjer Recursion* kelas (a, b, 0) ditunjukkan oleh persamaan di bawah ini.

$$f_s(x) = \frac{1}{1 - a\, f_X(0)} \sum_{y=1}^{x \wedge m} \left(a + \frac{b\, y}{x}\right) f_X(y)\, f_s(x-y)$$

Fungsi peluang kerugian agregat untuk data klaim asuransi kecelakaan PT Jasa Raharja *underwriting year* 2021 hingga 2022 adalah sebagai berikut:

$$f_s(x) = \frac{1}{1 - 2.142857\, f_X(0)} \sum_{y=1}^{x \wedge m} \left(a + \frac{(6{,}42408 \times 10^3)\, y}{x}\right) f_X(y)\, f_s(x-y)$$

Nilai ekspektasi dan variansi dari kerugian agregat klaim asuransi kecelakaan dapat dilihat pada Tabel 15.

Nilai variansi dan ekspektasi yang telah didapatkan akan digunakan pada perhitungan *risk premium*.

F. Menentukan Nilai Risk Premium

Perhitungan *risk premium* pada penelitian ini menggunakan tiga *premium principle* yaitu *expectation principle*, *standar deviation principle*, dan *dutch principle*. Sesuai dengan pembahasan pada persamaan 39 hingga persamaan 41, persamaan *risk premium* dapat dituliskan sebagai berikut:

1. *Expectation principle*

$$\Pi(X) = (1+m)(9{,}20588 \times 10^{11})$$

2. *Standar deviation principle*

$$\Pi(X) = 9{,}20588 \times 10^{11} + m\sqrt{1{,}03885 \times 10^{23}}$$

3. *Variance principle*

$$\Pi(X) 9{,}20588 \times 10^{11} + m(1{,}03885 \times 10^{23})$$

Nilai *risk premium* dipengaruhi oleh *m* atau *loading factor*, nilai *loading factor* bergantung pada masing-masing dari perusahaan asuransi. Pada penelitian kali ini, akan digunakan *loading factor* $m = 1$, $m = 1{,}5$, dan $m = 2$.

Dari Tabel 16 nilai *risk premium* tertinggi yaitu menggunakan perhitungan *variance principle*, sedangkan untuk perhitungan menggunakan *expectation principle* dan *standar deviation principle* tidak menghasilkan nilai yang jauh berbeda. Dapat diketahui pula bahwa semakin tingginya koefisien *loading factor* maka nilai *risk premium* juga semakin meningkat.

V. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, terdapat beberapa kesimpulan sebagai berikut: (1) Periode distribusi yang sesuai dalam memodelkan data klaim asuransi kecelakaan PT Jasa Raharja KPJR Mojokerto *underwriting year* 2021-2022 yaitu periode bulanan karena memiliki koefisien autokorelasi yang lebih rendah dibandingkan dengan periode harian maupun mingguan. Distribusi terbaik dalam memodelkan data frekuensi klaim yaitu distribusi

Poisson-Tweedie dengan nilai AIC yang lebih kecil yaitu sebesar 274,50. Nilai estimasi parameter dari distribusi *Poisson-Tweedie* adalah $\hat{a} = 2{,}14286, \hat{b} = 6{,}42408 \times 10^3, \hat{c} = 1{,}04907$. (2) Model kerugian agregat dibentuk dari distribusi frekuensi dan *severity* klaim yang telah terbentuk. Distribusi frekuensi data klaim mengikuti distribusi *Poisson-Tweedie* sedangkan distribusi *severity* data klaim mengikuti distribusi Lognormal dengan parameter$\mu = 22{,}147689$ dan $\sigma = 0{,}23259$. Dari parameter distribusi frekuensi dan *severity* tersebut kemudian terbetuk model kerugian agregat sebagai berikut: $f_s(x) = \frac{1}{1-2.142857\ f_X(0)} \sum_{y=1}^{x \wedge m} \left(2{,}14286 + \frac{(6{,}42408 \times 10^3)\ y}{x}\right) f_X(y)\ f_s(x-y)$. (3) Perhitungan *risk premium* dilakukan dengan tiga *premium principle* yang berbeda. Nilai *loading factor* (m) yang berbeda menghasilkan hasil yang semakin tinggi. Pada *expectation principle* menghasilkan nilai *risk premium* dengan *range* $[1{,}84117 \times 10^{12};\ 2{,}76176 \times 10^{12}]$. Kemudian, untuk perhitungan menggunakan *standard deviation principle* menghasilkan nilai *risk premium* dengan *range* $[1{,}24289 \times 10^{12};\ 1{,}56521 \times 10^{12}]$. Perhitungan menggunakan *variance principle* menghasilkan nilai *risk premium* dengan *range* $[1{,}03885 \times 10^{23};\ 2{,}07769 \times 10^{23}]$.

DAFTAR PUSTAKA

[1] Fara Lukita Umul Amaliah, Emy Siswanah, and S. D. Miasary, “Analisis jumlah klaim agregasi berdistribusi negative binomial dan besar klaim berdistribusi discreate uniform dengan menggunakan metode konvolusi,” *Journal of Mathematics: Theory and Applications*, vol. 1, no. 2, pp. 50–56, Nov. 2022, doi: 10.31605/jomta.v4i2.2010.

[2] S. Chen, Z. Wang, and M. Kelly, “Aggregate loss model with poisson-tweedie frequency,” *Big Data and Information Analytics*, vol. 6, no. 0, pp. 56–73, 2021, doi: 10.3934/bdia.2021005.

[3] H. Fitrianti, “Menentukan Premi Murni Menggunakan Generalized Liniear Models dan Model Copula dalam Asuransi Pertanian,” Departemen Aktuaria, Institut Teknologi Bandung, Bandung, 2015.

[4] R. E. Sya’banni, “Masalah Asuransi Kesehatan Penentuan Risk Premium (Studi Kasus Pembedahan),” Departemen Matematika, Institut Teknologi Bandung, Bandung, 2010.

[5] M. Firdaus, *Ekonometrika: Suatu Pendekatan Aplikatif.* Jakarta: Bumi Aksara, 2021. ISBN: 9786233280471.

[6] T. Manurung and M. Mananohas, “Taksiran distribusi aggregate loss asuransi mobil menggunakan fast fourier transform (fft) dalam menentukan premi murni,” *d\'Cartesian: Jurnal Matematika dan Aplikasi*, vol. 5, no. 2, p. 63, Oct. 2016, doi: 10.35799/dc.5.2.2016.13843.

[7] S. A. Klugman, H. H. Panjer, and G. E. Willmot, *Loss Models: From Data to Decisions*, 4th ed. Canada: Wiley, 2012. ISBN: 9781118315323.

[8] R. Andaria, “Penentuan Risk Premium Asuransi Kesehatan Rawat Jalan untuk Jenis Kelamin Laki-Laki : Studi Kasus,” Departemen Matematika, Institut Teknologi Bandung, Bandung, 2010.