Xplore (2022), Vol. 11, No. 3, 215-225
DOI:https://doi.org/10.29244/xplore.v11i3.1018

Xplore: Journal of Statistics
Available at https://journal.stats.id/index.php/xplore
e-ISSN: 2655-2744

Perbandingan Kinerja Regresi Conway-Maxwell-Poisson dan Poisson-Tweedie dalam Mengatasi Overdispersi Melalui Data Simulasi*

Ahmad Rifai Nasution[1], Kusman Sadik[2‡], Akbar Rizki[2]

[1]Audit and Risk Development Program, PT Astra International Tbk, Indonesia
[2]Department of Statistics, IPB University, Indonesia
[‡]corresponding author: kusmans@apps.ipb.ac.id

Abstract

Poisson regression is a standard method to model count data. Modeling count data frequently causes overdispersion which means that Poisson regression is less precise to model it as Poisson regression has the assumption of equidispersion. Overdispersion can be overcome by using Conway-Maxwell-Poisson (COM-Poisson) and Poisson Tweedie (Poisson-Tw) regression. The best model is determined based on the lowest value of RMSE, absolute bias, variance of parameter estimator, AIC, and BIC. This research uses simulation data. The response variable of simulation data is generated to follow Generalized Poisson distribution with combinations of ϕ and n. The result of simulation study shows that COM-Poisson and Compound Poisson-Tw are the alternatives to model overdispersed count data, but COM-Poisson is better to overcome overdispersion with higher dispersion parameter.

Keywords: COM-Poisson, Compound Poisson-Tw, overdispersion, Poisson.

1. Pendahuluan

Analisis regresi merupakan salah satu metode statistika yang digunakan untuk mengetahui hubungan antara satu peubah respon dengan satu atau lebih peubah penjelas. Secara umum, analisis regresi yang sering digunakan adalah regresi klasik dengan peubah responnya merupakan data kontinu yang berdistribusi sebaran normal. Namun demikian, peubah respon berupa data cacah dan tidak menyebar normal sering dijumpai sehingga regresi klasik ini kurang tepat dalam memodelkan data tersebut (Agresti 2018).

Generalized Linear Model (GLM) merupakan model pengembangan dari model regresi klasik untuk mengatasi peubah respon yang tidak menyebar normal. Peubah

* Received: Aug 2022 ; Reviewed: Aug 2022 ; Published: Sep 2022

respon dalam GLM diasumsikan memiliki sebaran yang termasuk dalam sebaran keluarga Eksponensial. GLM terdiri atas tiga komponen utama, yaitu komponen acak, komponen sistematik, dan fungsi penghubung (Agresti 2018). Salah satu sebaran yang termasuk ke dalam keluarga sebaran Eksponensial adalah sebaran Poisson. Sebaran ini merupakan salah satu dari sebaran diskrit yang menyatakan jumlah kejadian yang terjadi dalam selang waktu tertentu. Pendekatan dengan model regresi Poisson sering digunakan untuk memodelkan peubah respon berupa data cacah dengan peubah penjelas berupa data diskrit atau cacah, kontinu, kategorik, dan data campuran.

Regresi Poisson merupakan analisis regresi nonlinier dalam memodelkan data dengan peubah respon berupa data cacah dan dimodelkan menggunakan GLM. Terdapat asumsi yang harus dipenuhi dalam model regresi Poisson, yaitu equidispersi. Equidispersi berarti nilai ragam peubah respon dan rataannya sama. Pada kasus nyata, data cacah sering memperlihatkan ragam yang lebih besar dari rataannya (*overdispersi*) atau ragam lebih kecil dari rataannya (*underdispersi*) (Sellers dan Premeaux 2020). Menurut Hilbe (2011), overdispersi sering terjadi dalam pemodelan data cacah karena adanya keragaman dalam peubah respon dan adanya korelasi positif antar peubah respon. Korelasi antar peubah respon ini dapat terjadi dalam analisis multivariat atau peubah ganda, yaitu analisis yang mempertimbangkan lebih dari satu peubah respon dalam pemodelannya. Selain itu, overdispersi juga disebabkan oleh adanya pencilan pada data dan pengamatan yang hilang pada peubah penjelas (Hardin dan Hilbe 2018). Overdispersi mengakibatkan galat baku dari parameter menjadi bias ke bawah dan signifikansi dari peubah penjelas menjadi bias ke atas, sehingga kesimpulan yang didapatkan tidak valid. Kesimpulan yang tidak valid ini mengindikasikan bahwa peubah penjelas cenderung dianggap berpengaruh tetapi pada kenyataannya peubah penjelas tersebut belum tentu berpengaruh.

Sellers dan Premeaux (2020) memberikan alternatif pendekatan untuk memodelkan data cacah dengan kasus overdispersi maupun underdispersi, yaitu regresi Conway-Maxwell-Poisson (COM-Poisson). Model ini memiliki fleksibilitas dalam memodelkan data dengan berbagai tingkatan dispersi. Hal ini sejalan dengan penelitin Hayati *et al.* (2018) yang menghasilkan simpulan bahwa model regresi COM-Poisson lebih baik dan fleksibel dalam mengatasi *under* dan *over*-dispersi pada data cacah jika dibandingkan model regresi *Generalized* Poisson dan Binomial Negatif. Selain itu, penelitian dilakukan oleh Debrabant *et al.* (2018) dalam memodelkan jumlah kasus kecelakaan lalu lintas dengan data overdispersi. Hasil penelitian tersebut menunjukkan bahwa regresi Poisson-Tweedie (Poisson-Tw) memberikan kinerja yang baik. Saha *et al.* (2020) juga melakukan pemodelan data kasus kecelakaan dengan beberapa model dari keluarga sebaran Tweedie. Hasil penelitiannya menunjukkan bahwa model *Compound* Poisson-Tweedie lebih baik jika dibandingkan model dari keluarga sebaran Tweedie lainnya dengan parameter dispersi konstan. Oleh karena itu, Kajian data simulasi dilakukan untuk membandingkan regresi Poisson sebagai model dasar, COM-Poisson, dan *Compound* Poisson-Tw untuk memodelkan data dengan kasus overdispersi. Kriteria model terbaik ditentukan berdasarkan nilai terkecil dari *Root Mean Square Error* (RMSE), *absolute bias*, ragam penduga parameter, *Akaike Information Criterion*

(AIC), dan *Bayesian Information Criterion* (BIC).

2. Metodologi

2.1 Data

Data simulasi digunakan untuk mengevaluasi kebaikan model regresi COM-Poisson dan *Compound* Poisson-Tweedie dalam mengatasi kasus overdispersi pada data cacah. Regresi Poisson juga digunakan sebagai model dasar dalam pemodelan data cacah. Terdapat satu peubah respon (Y) dan tiga peubah penjelas (X_1, X_2, X_3) yang digunakan pada kajian data simulasi. Simulasi dalam penelitian ini dilakukan dengan beberapa penyesuaian dari algoritme Yang *et al.* (2009) yang dalam penelitiannya hanya menggunakan satu peubah penjelas berdistribusi sebaran seragam (0, 1) dengan $n = 20; 50; 100;$ $\beta_0 = 2;$ $\beta_1 = -0{,}5;$ $\phi = [0; 0{,}4]$, dan pengulangan yang dilakukan 10.000 kali. Data simulasi pada penelitian ini didasarkan pada model *Generalized* Poisson (GP) dengan model yang dibentuk dapat dirumuskan seperti pada persamaan (1).

$$\ln(\mathrm{y}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \tag{1}$$

Nilai parameter pada persamaan (1) ditentukan oleh peneliti, yaitu $\beta_0 = 0{,}2$; $\beta_1 = 0{,}3$; $\beta_2 = 0{,}5$; dan $\beta_3 = -0{,}6$. Pemilihan parameter tersebut mewakili parameter yang bertanda positif dan negatif, serta lebih bervariasi jika dibandingkan dengan simulasi yang dilakukan oleh Yang *et al.* (2009) sebelumnya. Peubah penjelas X_1, X_2, X_3 dibangkitkan mengikuti sebaran normal yang rata-ratanya sebesar 2; 0,2; dan 0,5 dengan ragam satu. Pemilihan parameter pada peubah-peubah penjelas tersebut dimaksudkan untuk kemudahan komputasi dalam pembangkitan data. Ukuran contoh yang digunakan adalah $n = 10; 30; 100;$ dan 500. Ukuran contoh yang ditetapkan mewakili ukuran contoh kecil, sedang, dan besar. Parameter dispersi ϕ dari sebaran GP yang digunakan yaitu 0; 0,1; 0,25; 0,45; dan 0,7. Parameter dispersi ini merepresentasikan kondisi equidispersi $(\phi = 0)$ dan overdispersi pada berbagai tingkatan dispersi (kecil, sedang, tinggi). Evaluasi model dilakukan berdasarkan RMSE, *absolute* bias, ragam penduga parameter, AIC, dan BIC. Simulasi dalam penelitian ini dilakukan dengan 500 kali pengulangan untuk memperoleh hasil yang stabil.

2.2 Prosedur analisis Data

Prosedur analisis data simulasi dapat dilihat pada Gambar 1 and secara umum dibagi menjadi empat tahapan, yaitu pembangkitan data, pemodelan data, identifikasi overdispersi, dan mengevaluasi model. Prosedur analisis data secara lebih rinci dijelaskan sebagai berikut:

1. Pembangkitan data
   a. Membangkitkan peubah X_1, X_2, X_3 sebanyak n yang mengikuti sebaran normal dengan rata-rata $2; 0{,}2; 0{,}5$ dan ragam bernilai satu. Ketiga peubah penjelas diasumsikan tidak berkorelasi kuat.
   b. Menghitung nilai μ_i berdasarkan persamaan (2) dengan $\beta_0, \beta_1, \beta_2, \beta_3$ yang telah ditetapkan sebelumnya.

$$\mu_i = \exp\left(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3}\right) \qquad (2)$$

c. Membangkitkan Y_i sebanyak n yang mengikuti sebaran GP $[Y_i \sim GP(\mu_i, \phi)]$ dengan parameter dispersi $\phi = 0; 0{,}1; 0{,}25; 0{,}45$; dan 0,7.

2. Pemodelan data
Pemodelan pada data hasil simulasi dilakukan dengan menggunakan regresi Poisson sebagai model dasar, COM-Poisson, dan *Compound* Poisson-Tweedie. Model regresi yang dibentuk dapat dirumuskan pada persamaan (3) (Hardin dan Hilber 2018).

$$y_i = \exp\left(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}\right) \qquad (3)$$

Keterangan:
i = 1, 2, 3, …, n
n = banyaknya amatan
y_i = nilai peubah respon pada amatan ke-i
β_0 = konstanta
β_k = koefisien regresi pada peubah penjelas ke-k
k = jumlah peubah penjelas di dalam model

3. Identifikasi overdispersi
Identifikasi overdispersi pada data hasil simulasi dilakukan dengan menggunakan nilai rasio *deviance* dengan derajat bebasnya dan uji skor.
a. Nilai rasio didapatkan dengan menghitung nilai *deviance* pada regresi Poisson dan membagi nilai tersebut dengan derajat bebasnya. Jika nilai rasio *deviance* dengan derajat bebasnya lebih besar dari satu maka terjadi overdispersi (Ulfa *et al.* 2021). Nilai *deviance* dapat dirumuskan pada persamaan (4).

$$D = 2\sum_{i=1}^{n} y_i \ln\frac{y_i}{\hat{y}_i} - (y_i - \hat{y}_i) \qquad (4)$$

dengan y_i adalah nilai amatan ke-i dari peubah respon, $\hat{y}_i$ merupakan nilai dugaan peubah respon pada amatan ke-i dengan $i = 1, 2, \ldots, n$. Jumlah amatan dinotasikan dengan n.
b. Melakukan uji skor dengan hipotesis yang digunakan adalah $H_0: \phi = 0$ dan $H_1: \phi > 0$ dengan taraf nyata 0,05. Uji skor digunakan karena dianggap relatif lebih baik dari uji lainnya (Payne *et al.* 2018). Statistik uji yang digunakan adalah:

$$S_\phi = \frac{\sum_{i=1}^{n}[(y_i - \hat{y}_i)^2 - y_i]}{\sqrt{2\sum_{i=1}^{n}\hat{y}_i^2}} \qquad (5)$$

dengan $\hat{y}_i$ nilai dugaan peubah respon pada amatan ke-i dari model Poisson dan y_i adalah nilai amatan ke-i dari peubah respon. Penolakan H_0 terjadi jika nilai statistik uji (S_ϕ) lebih besar dari nilai Z_α atau *p-value* lebih kecil dari taraf nyata 0,05 sehingga dapat dikatakan terjadi overdispersi.

4. Evaluasi model
Evaluasi model pada data simulasi dilakukan menggunakan nilai RMSE, *absolute* bias, ragam penduga parameter, AIC, dan BIC. Nilai RMSE, *absolute* bias, nilai ragam penduga parameter, AIC, dan BIC yang semakin kecil, maka model semakin baik (Rajitha dan Sakthivel 2019). Nilai kebaikan model ini dapat dirumuskan sebagai berikut:

a. Formula perhitungan RMSE didasarkan pada hasil prediksi untuk masing-masing model.

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}} \quad (6)$$

dengan n adalah banyaknya jumlah amatan, y_i adalah nilai peubah respon pada amatan ke-i, $\hat{y}_i$ adalah nilai dugaan peubah respon pada amatan ke-i.

b. Formula untuk menghitung nilai *absolute* bias didasarkan pada perbedaan antara penduga parameter dengan parameter sebenarnya. Nilai *absolute* bias digunakan untuk melihat kinerja hasil pendugaan masing-masing model.

$$Bias(\hat{\beta}_j) = |\beta_j - \hat{\beta}_j| \quad (7)$$

dengan $\hat{\beta}_j$ adalah dugaan parameter regresi ke-j dan β_j adalah nilai parameter regresi ke-j sebenarnya untuk $j = 0, 1, 2, 3$.

c. Formula nilai ragam penduga parameter didasarkan pada keragaman dugaan parameter yang didapatkan dari tiap ulangan pada data simulasi. Nilai ragam penduga yang semakin kecil mengindikasikan bahwa model semakin baik dan stabil.

$$S_b^2 = \frac{\sum_{i=1}^{m} (b_i - \bar{b})^2}{m-1} \quad (8)$$

dengan b_i adalah penduga parameter pada ulangan ke-i, $\bar{b}$ merupakan rata-rata penduga parameter, dan m merupakan banyaknya ulangan.

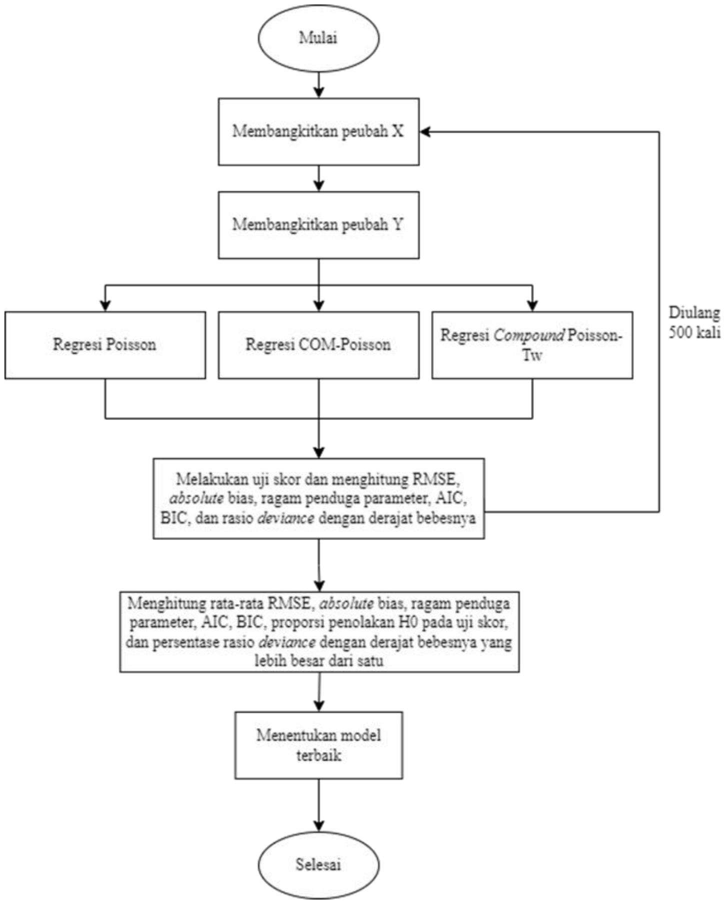
d. Formula perhitungan AIC dan BIC didasarkan pada persamaan (9) dan (10).

$$AIC = -2 \ln L + 2p \quad (9)$$

$$BIC = -2 \ln L + p \ln(n) \quad (10)$$

dengan n adalah banyaknya jumlah amatan, $\ln L$ merupakan nilai maksimum dari fungsi *log-likelihood,* serta p adalah banyaknya parameter yang diduga dalam model.

Penentuan model terbaik dilakukan dengan menghitung rata-rata nilai RMSE, *absolute* bias, AIC, dan BIC, serta perhitungan ragam penduga parameter pada setiap kombinasi parameter dan jumlah amatan yang sudah ditetapkan.



3. Hasil dan Pembahasan

3.1 Identifikasi Overdisperi

Overdispersi pada data amatan dapat diidentifikasi dengan nilai rasio *deviance* dengan derajat bebasnya atau menggunakan uji skor. Rasio *deviance* dengan derajat bebasnya ini sering digunakan dalam beberapa penelitian untuk mendeteksi overdisperi karena tidak memerlukan pengujian hipotesis dalam pengambilan keputusannya. Suatu data dapat dikatakan mengalami overdispersi jika nilai rasio *deviance* dengan derajat bebasnya lebih besar dari satu.

Tabel 1 menyajikan persentase rasio antara *deviance* pada regresi Poisson dengan derajat bebasnya. Persentase ini menunjukkan banyaknya rasio yang bernilai lebih besar dari satu dalam 500 pengulangan yang dilakukan dalam simulasi. Nilai rasio yang lebih besar dari satu mengindikasikan overdispersi. Nilai ϕ yang semakin besar maka persentase rasio yang lebih besar dari satu semakin besar sehingga semakin banyak data yang mengalami overdispersi. Hal ini berkaitan bahwa ϕ sendiri yang merupakan parameter dispersi. Selain itu, setiap kenaikan n dan ϕ juga memperlihatkan persentase nilai rasio yang semakin besar. Contoh dapat dilihat pada nilai $\phi = 0{,}45$. Semua data simulasi yang dibangkitkan dengan 500 kali ulangan sudah diidentifikasi mengalami overdispersi pada $n = 30$. Namun demikian, persentase rasio ini cukup tinggi atau lebih dari 50 persen data diidentifikasi mengalami overdispersi pada $\phi = 0$. Persentase rasio ini juga semakin meningkat ketika n semakin besar sehingga menggunakan nilai rasio *deviance* dengan derajat bebasnya pada kondisi data tidak mengalami overdispersi kurang disarankan karena data tersebut akan cenderung dianggap mengalami overdispersi padahal sebenarnya data tidak overdispersi.

Tabel 1 Persentase rasio *deviance* dengan derajat bebasnya pada regresi Poisson yang lebih besar dari satu pada berbagai kondisi n dan $\boldsymbol{\phi}$

ϕ	n			
	10	30	100	500
0,00	56,20	62,80	80,00	95,80
0,10	67,80	82,00	98,00	100,00
0,25	81,80	97,00	100,00	100,00
0,45	92,80	100,00	100,00	100,00
0,70	98,40	100,00	100,00	100,00

Data overdispersi pada amatan juga dapat diidentifikasi dengan menggunakan statistik uji skor. Uji skor menggunakan hipotesis berikut: $H_0: \phi = 0; H_1: \phi > 0$. Hipotesis nol ditolak pada saat statistik uji skor lebih besar dari nilai Z_α sehingga mengindikasikan bahwa nilai ϕ pada data lebih besar dari nol atau data mengalami overdispersi. Hasil uji skor pada penelitian ini direpresentasikan dengan proporsi penolakan H_0 dari 500 kali ulangan. Semakin besar proporsi penolakan H_0 dapat juga mengindikasikan bahwa kuasa uji semakin besar sehingga kinerja pengujian semakin baik. Kuasa uji merupakan peluang menolak hipotesis nol yang salah dalam suatu pengujian (Park 2015). Penelitian ini menggunakan hipotesis nol ketika $\phi = 0$ atau data tidak mengalami overdispersi. Hipotesis nol tersebut mengindikasikan bahwa kuasa uji sebagai peluang untuk menolak hipotesis ketika ϕ yang digunakan tidak bernilai nol.

Tabel 2 menginformasikan proporsi penolakan H_0 pada uji skor. Setiap jumlah amatan $n = 10; 30; 100; 500$ menunjukkan bahwa semakin besar nilai ϕ yang digunakan untuk membangkitkan data simulasi, maka proporsi penolakan H_0 yang didapatkan semakin besar. Proporsi penolakan H_0 yang dihasilkan juga meningkat seiring dengan semakin besarnya parameter dispersi yang digunakan. Hasil simulasi untuk data equidispersi ($\phi = 0$) menunjukkan proporsi penolakan H_0 yang didapatkan untuk masing-masing n sangat kecil karena banyak hipotesis nol yang tidak ditolak pada uji skor untuk 500 kali ulangan yang dilakukan sehingga dapat dikatakan data tidak mengalami overdispersi. Oleh karena itu, uji skor lebih disarankan penggunaannya ketika nilai ϕ adalah nol atau mendekati nol.

Kajian data simulasi yang mengalami overdispersi ($\phi > 0$) pada Tabel 2 memperlihatkan bahwa proporsi penolakan H_0 dari masing-masing n mulai meningkat signifikan jika dibandingkan pada saat kondisi data tanpa overdispersi ($\phi = 0$). Semakin besar nilai ϕ yang digunakan, maka jarak antara hipotesis nol dan hipotesis alternatif juga semakin besar sehingga hasil uji skor cenderung akan menolak hipotesis nol. Selain itu, kajian simulasi pada $\phi \geq 0{,}25$ menunjukkan bahwa semakin besar n yang digunakan, maka kinerja uji skor semakin baik. Hal ini dapat dilihat ketika $\phi = 0{,}25$ dengan $n = 500$, proporsi penolakan H_0 yang dihasilkan sebesar satu sehingga dapat dikatakan bahwa semua data dianggap mengalami overdispersi oleh uji skor. Hal yang yang sama juga dapat dilihat ketika $\phi = 0{,}45$ dan $0{,}7$ dengan $n = 100; 500$, yaitu proporsi penolakan H_0 yang didapatkan sebesar satu.

Tabel 2 Proporsi menolak H_0 pada uji skor dengan 500 kali ulangan

ϕ	n			
	10	30	100	500
0,00	0,002	0,004	0,032	0,036
0,10	0,008	0,026	0,202	0,708
0,25	0,042	0,124	0,844	1,000
0,45	0,222	0,640	1,000	1,000
0,70	0,580	0,988	1,000	1,000

3.2 Evaluasi Model

Mengkaji dan mengevaluasi kinerja model dalam memodelkan data dengan berbagai tingkatan dispersi dapat dilakukan dengan melihat nilai RMSE dari model tersebut. Besaran RMSE digunakan untuk mengukur kesalahan dalam proses prediksi. Model yang baik dalam memprediksi ditunjukkan dengan nilai RMSE dan *Mean Square Error* (MSE) yang kecil. RMSE adalah nilai akar kuadrat dari MSE yang digunakan untuk mengukur kesalahan suatu model (Hastomo *et al.* 2021).

Tabel 3 Rata-rata RMSE setiap model pada berbagai kondisi n dan ϕ

n	ϕ	RMSE		
		Poisson	COM-Poisson	*Compound* Poisson-Tw
10	0,00	0,818	0,998	0,803
	0,10	0,886	0,999	0,867
	0,25	1,019	1,005	0,988
	0,45	1,273	1,009	1,179
	0,70	1,952	1,022	1,533
30	0,00	0,973	1,006	0,972
	0,10	1,056	1,016	1,053
	0,25	1,224	1,034	1,208
	0,45	1,568	1,075	1,453
	0,70	2,530	1,170	1,852
100	0,00	1,034	1,069	1,032
	0,10	1,126	1,011	1,122
	0,25	1,306	1,035	1,288
	0,45	1,680	1,084	1,515
	0,70	2,739	1,223	1,916
500	0,00	1,044	1,682	1,043
	0,10	1,137	1,016	1,134
	0,25	1,317	1,051	1,201
	0,45	1,692	1,128	1,536
	0,70	2,755	1,357	1,970

Tabel 3 menunjukkan nilai rata-rata RMSE dari model regresi Poisson, COM-Poisson, dan *Compound* Poisson-Tw cenderung meningkat seiring bertambahnya jumlah amatan. Proses simulasi pada $\phi = 0$ memperlihatkan bahwa rata-rata RMSE yang dihasilkan regresi *Compound* Poisson-Tw cenderung lebih kecil untuk semua jumlah amatan tetapi perbedaannya tidak terlalu signifikan bahkan sangat kecil jika dibandingkan rata-rata RMSE regresi Poisson.

Kajian simulasi untuk $\phi > 0$ pada Tabel 3 menginformasikan bahwa rata-rata RMSE regresi *Compound* Poisson-Tw lebih kecil jika dibandingkan regresi Poisson untuk setiap jumlah amatan. Selain itu, regresi COM-Poisson dianggap lebih baik dalam memodelkan data overdispersi karena cenderung menghasilkan rata-rata RMSE yang lebih kecil. Selisih perbedaan rata-rata RMSE COM-Poisson dan *Compound* Poisson-Tw semakin besar dengan bertambahnya nilai ϕ. Hal ini berarti bahwa regresi COM-Poisson semakin baik kinerjanya seiring dengan bertambahnya nilai ϕ.

Evaluasi selanjutnya dilakukan terhadap nilai *absolute* bias dan ragam penduga parameter. Tabel 4 menunjukkan rata-rata nilai *absolute* bias dan ragam penduga bagi parameter pada regresi Poisson, COM-Poisson, dan *Compound* Poisson-Tw untuk $n = 100$. Rata-rata *absolute* bias dan ragam penduga parameter cenderung tidak berbeda jauh antar ketiga model regresi tersebut tetapi masih terdapat perbedaannya. Regresi Poisson dan *Compound* Poisson-Tw cenderung memiliki rata-rata *absolute* bias dan ragam penduga yang kecil saat $\phi = 0$. Regresi COM-Poisson dan *Compound* Poisson-Tw cenderung memiliki kinerja yang sama dengan nilai *absolute* bias yang lebih kecil dari regresi Poisson pada data overdispersi. Pola yang sama juga terlihat pada nilai ragam penduga parameter.

Tabel 4 Rata-rata *absolute* bias dan ragam penduga parameter pada $\boldsymbol{n = 100}$

ϕ	Parameter	*Absolute* bias			Ragam penduga parameter		
		Pois	COM-Pois	Comp Pois-Tw	Pois	COM-Pois	Comp Pois-Tw
0,00	β_0	0,103	0,103	0,103	0,016	0,016	0,016
	β_1	0,035	0,036	0,035	0,002	0,005	0,002
	β_2	0,046	0,047	0,046	0,003	0,003	0,003
	β_3	0,059	0,061	0,059	0,005	0,005	0,005
0,10	β_0	0,135	0,133	0,133	0,021	0,018	0,016
	β_1	0,038	0,037	0,037	0,009	0,002	0,002
	β_2	0,052	0,049	0,049	0,004	0,003	0,003
	β_3	0,065	0,062	0,062	0,006	0,004	0,003
0,25	β_0	0,275	0,274	0,274	0,022	0,022	0,020
	β_1	0,042	0,040	0,041	0,004	0,002	0,002
	β_2	0,054	0,053	0,052	0,005	0,004	0,004
	β_3	0,068	0,066	0,066	0,007	0,005	0,005
0,45	β_0	0,580	0,578	0,579	0,030	0,029	0,029
	β_1	0,060	0,048	0,051	0,004	0,003	0,003
	β_2	0,064	0,062	0,063	0,009	0,006	0,006
	β_3	0,083	0,078	0,078	0,010	0,008	0,009
0,70	β_0	1,174	1,158	1,167	0,057	0,051	0,055
	β_1	0,068	0,064	0,064	0,006	0,003	0,006
	β_2	0,090	0,085	0,086	0,012	0,008	0,011
	β_3	0,105	0,102	0,103	0,018	0,016	0,017

Nilai AIC dan BIC juga dapat digunakan untuk mengukur kinerja kebaikan model. Kedua nilai ini didasarkan pada fungsi *log-likelihood*. Rata-rata AIC dan BIC untuk setiap model disajikan pada Tabel 5. Tabel 5 menginformasikan bahwa regresi *Compound* Poisson-Tw menghasilkan rata-rata AIC dan BIC terkecil saat kondisi data equidispersi. Hal ini berarti regresi *Compound* Poisson-Tw baik digunakan untuk memodelkan data tersebut. Namun demikian, rata-rata AIC dan BIC regresi Poisson juga tidak terlalu besar sehingga model ini masih cukup baik digunakan untuk memodelkan data equidispersi.

Kajian simulasi untuk data overdispersi pada Tabel 5 menunjukkan bahwa regresi *Compound* Poisson-Tw memiliki kinerja yang baik karena menghasilkan rata-rata AIC dan BIC terkecil. Walaupun demikian, regresi COM-Poisson menghasilkan kinerja yang lebih baik pada tingkatan dispersi yang tinggi. Model regresi COM-Poisson akhirnya memiliki nilai AIC dan BIC terkecil pada $\phi = 0{,}7$.

Tabel 5 Rata-rata AIC dan BIC setiap model pada berbagai kondisi n dan ϕ

n	ϕ	AIC			BIC		
		Poisson	COM-Poisson	*Comp* Poisson-Tw	Poisson	COM-Poisson	*Comp* Poisson-Tw
10	0,00	38,147	39,216	27.533	39,357	40,729	28.744
	0,10	41,083	40,884	30,457	42,293	42,398	31,668
	0,25	44,993	45,251	35,192	46,203	46,764	36,403
	0,45	53,372	51,849	44,149	54,582	53,362	45,359
	0,70	82,306	63,538	63.543	83,516	64,091	64.753
30	0,00	109,228	109,785	71,288	114,833	116,791	76,893
	0,10	116,385	117,220	79,902	121,990	124,226	85,507
	0,25	131,509	130,148	101,451	137,173	137,154	107,055
	0,45	166,862	150,757	141,581	172,477	157,763	147,187
	0,70	300,945	187,532	188,331	306,550	193,539	194.936
100	0,00	346,621	347,634	200,775	357,042	360,660	211,196
	0,10	373,176	372,847	234,122	383,596	381,872	244,542
	0,25	428,272	415,069	319,681	438,693	428,095	330,101
	0,45	559,433	483,154	479,216	569,854	496,181	489,637
	0,70	1066,471	609,146	612,091	1076,892	618,669	622.511
500	0,00	1639,273	1708,359	917,189	1656,131	1729,432	934,047
	0,10	1771,340	1763,029	1080,371	1788,199	1774.102	1097,230
	0,25	2045,731	1968,420	1449,244	2062,589	1989,493	1466,103
	0,45	2698,687	2374,624	2303,168	2715,546	2395,697	2320,027
	0,70	5226.899	2875,259	2927.344	5243,757	2896,043	2944.202

4. Simpulan

Kajian pemodelan dengan menggunakan data simulasi menunjukkan bahwa regresi Poisson, COM-Poisson, dan *Compound* Poisson memiliki kinerja yang stabil dan konsisten dalam pendugaan parameter regresi. Regresi *Compound* Poisson-Tw dan regresi Poisson memiliki kinerja yang baik dalam memodelkan data equidispersi. Walaupun demikian, regresi Poisson lebih disarankan karena mempertimbangkan kesederhanaannya dalam memodelkan data. Kajian pada data overdispersi menunjukkan bahwa regresi *Compound* Poisson-Tw baik dalam memodelkan data overdispersi untuk tingkatan dispersi yang tidak terlalu tinggi, sedangkan regresi COM Poisson menunjukkan kinerja yang baik pada data dengan tingkatan dispersi yang tinggi atau lebih dari 0,45.

Daftar Pustaka

Agresti A. 2018. *Introduction to Categorical Data Analysis, Second Edition*. New Jersey (US): John Wiley and Sons, Inc.

Debrabant B, Halekoh U, Bonat WH, Hansen DL, Hjelmborg J, Lauritsen J. 2018. Identifying traffic accident black sports with Poisson-Tweedie models. *Accidental Analysis and Prevention*. 111: 147-154.

Hardin JW, Hilbe JM. 2018. *Generalized Linear Models and Extensions 4th Edition*. USA: Stata Press.

Hastomo W, Karno ASB, Kalbuana N, Nisfiani E, Lussiana ETP. 2021. Optimasi deep learning untuk prediksi sahan di masa pandemi covid-19. *Jurnal Edukasi dan Penelitian Informatika*. 7 (2): 133-140.

Hayati M, Sadik K, Kurnia A. 2018. Conway-maxwell poisson distribution: approach for over-and under-dispersed count data modelling. *IOP Conference Series: Earth and Environmental Sciences*. 187 (1).

Hilbe JM. 2011. *Negative Binomial Regression Second Edition*. Cambridge (UK): Cambridge University Press.

Park HM. 2015. *Hypothesis Testing and Statistical Power of a Test*. Bloomington (US): Indiana University Pr.

Payne EH, Gebregziabher M, Hardin J, Ramakrishnan V. 2018. An empirical approach to determine a threshold for assessing overdispersion in Poisson and negative binomial models for count data. *Communication in Statistics*. 47 (5):1-17.

Rajitha CS, Sakthivel KM. 2019. Model selection for count data with excess number of zero counts. *American Journal of Applied Mathematics and Statistics*. 7 (1): 43-51.

Saha D, Alluri P, Dumbaugh E, Gan A. 2020. Application of the Poisson-Tweedie distribution in analyzing crash frequency data. *Accident Analysis and Prevention*. 137:105.

Sellers KF, Premeaux B. 2020. Conway-maxwell-poisson regression models for dispersed count data. *WIREs Computational Statistics*. 13 (6).

Ulfa YA, Soleh AM, Sartono B. Handling overdispersion in the Poisson regression model with Negative Binomial for the number of new cases of leprosy in Java. *Indonesian Journal of Statistics and Its Applications*. 5 (1): 1-13.

Yang Z, Hardin JW, Addy CL. 2009. A score test for overdispersion in poisson regression based on generalized poisson-2 model. *JSPI.* 139: 1541-1521.