

Bayesian-Negative Binomial Regression on Underreported Counts of Indonesian Female Trafficking Cases in 2023 using MCMC

Bayesian-Regresi Binomial Negatif Untuk *Underreported Counts* pada Kasus Perdagangan Perempuan di Indonesia Tahun 2023 menggunakan MCMC

Nabilah Disya Arviani¹, Reny Rian Marlina^{*2}

^{1,2}*Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung*

Email: ¹*ndisyaaa@gmail.com*, ^{*2}*renyrianmarlana@unisba.ac.id*

**Corresponding Author*

Received: 24 February 2025, revised: 26 March 2025, accepted: 29 March 2025

Abstract

The five-year report of the Task Force for the Prevention and Handling of Human Trafficking shows that in Indonesia from 2015 to 2019, 87.58% of human trafficking victims were women. Data of human trafficking cases often suffer from underreporting counts, where the number of reported incidents is smaller than the actual number of incidents. This study aims to estimate the actual number of cases of Indonesian female trafficking in 2023 using negative binomial regression underreported counts. Parameter estimation is conducted using a Bayesian approach through the Markov Chain Monte Carlo (MCMC) method with Gibbs sampling algorithm with a total of 5000 iterations and a burn-in period after 3000 iterations. This study utilizes secondary data sourced from Central Java Data Portal and Central Bureau of Statistics Indonesia, with the dependent variable is the reported number of Indonesian female trafficking in 2023 (Y) and three independent variables that are factors influencing female trafficking: the percentage of the poor population (X_1), the open unemployment rate (X_2), and the level of high school or equivalent education completion (X_3). The estimated actual number of female trafficking cases in 2023 in Aceh and West Sumatra is 4 and 2 cases respectively, with the unreported number of female trafficking cases amounting to 2 cases each. The average of actual number of Indonesian female trafficking in 2023 is 9 cases, while the average of unreported number of Indonesian female trafficking in 2023 is 2 cases.

Keywords: Bayesian, Negative Binomial Regression, Underreported Counts, Women's Trafficking.



Abstrak

Laporan lima tahunan Gugus Tugas Pencegahan dan Penanganan Tindak Pidana Perdagangan Orang (GTPP-TPPO) menunjukkan bahwa di Indonesia, pada tahun 2015-2019, sebanyak 87,58% korban perdagangan orang merupakan perempuan. Data kasus perdagangan orang sering kali mengalami *underreporting counts*, di mana jumlah kasus yang dilaporkan lebih kecil daripada jumlah kasus yang sebenarnya. Penelitian ini bertujuan untuk mengetahui jumlah kasus perdagangan perempuan yang sebenarnya di Indonesia tahun 2023 menggunakan regresi Binomial Negatif *underreported counts*. Penaksiran parameter model dilakukan menggunakan pendekatan Bayesian melalui metode *Markov Chain Monte Carlo* (MCMC) dengan algoritma *Gibbs sampling* dan iterasi sebanyak 5000 iterasi serta *burn-in period* setelah 3000 iterasi. Data yang digunakan dalam penelitian ini berasal dari *website* portal data Jawa Tengah dan Badan Pusat Statistik Indonesia dengan variabel dependen yaitu banyaknya kasus perdagangan perempuan di Indonesia tahun 2023 (Y) serta 3 variabel independen berupa faktor-faktor yang mempengaruhi perdagangan perempuan yaitu persentase penduduk miskin (X_1), tingkat pengangguran terbuka (X_2), dan tingkat penyelesaian pendidikan jenjang SMA/ sederajat (X_3). Taksiran banyaknya kasus perdagangan Perempuan yang sebenarnya pada tahun 2023 di Provinsi Aceh dan Sumatera Barat secara berturut-turut yaitu 4 dan 2 kasus dengan kasus perdagangan perempuan yang tidak dilaporkan sebanyak masing-masing 2 kasus. Rata-rata banyaknya kasus perdagangan perempuan yang sebenarnya di Indonesia pada tahun 2023 yaitu sebanyak 9 kasus, sedangkan rata-rata banyaknya kasus perdagangan perempuan yang tidak dilaporkan di Indonesia pada tahun 2023 yaitu sebanyak 2 kasus.

Kata kunci: Bayesian, Perdagangan Perempuan, Regresi Binomial Negatif, *Underreported Counts*.

1. PENDAHULUAN

Data yang diperoleh dari hasil pelaporan dapat mengalami *misreporting counts* jika laporan yang diperoleh mengenai banyaknya kejadian yang dilaporkan berbeda dengan banyaknya kejadian yang sebenarnya [6]. *Misreporting counts* dalam data pelaporan dibedakan menjadi dua jenis, yaitu *underreporting counts* dan *overreporting counts*. *Underreporting counts* terjadi ketika jumlah kejadian yang dilaporkan lebih rendah dari jumlah kejadian yang sebenarnya, sedangkan *overreporting counts* terjadi ketika jumlah kejadian yang dilaporkan lebih tinggi dari jumlah kejadian yang sebenarnya.

Underreported counts terjadi karena terdapat permasalahan dalam pencatatan dan pengumpulan data mengenai banyaknya kejadian yang diamati yang disebabkan oleh faktor tertentu [5]. *Underreported counts* dapat terjadi pada kasus perdagangan orang, di mana perdagangan orang sering kali dilakukan secara diam-diam sehingga mengakibatkan korban enggan melapor serta terjadi kesalahan dalam pencatatan korban perdagangan orang yang tercatat sebagai orang hilang. Perdagangan orang dapat terjadi karena faktor-faktor tertentu seperti faktor ekonomi, sedikitnya lapangan pekerjaan yang tersedia [7], dan tingkat pendidikan [1].

Laporan lima tahunan dari Gugus Tugas Pencegahan dan Penanganan Tindak Pidana Perdagangan Orang (GTPP-TPPO) menyebutkan bahwa di Indonesia terdapat sebanyak 2648 korban perdagangan orang yang terdiri atas 2319 perempuan dan 329 laki-laki pada tahun 2015-2019. Angka tersebut menunjukkan bahwa pada tahun 2015-2019 sebanyak 87,58% korban perdagangan orang di Indonesia merupakan perempuan. Mengacu pada data yang diperoleh dari Sistem Informasi *Online* Perlindungan Perempuan dan Anak (Simfoni PPA) tentang jumlah kasus perdagangan orang di Indonesia, pada tahun 2020 terdapat sebanyak 442 kasus, di mana angka tersebut menunjukkan kenaikan yang paling besar selama 6 tahun terakhir dari tahun 2019 yaitu sebesar 86,73%. Kemudian tahun 2021 mencapai jumlah kasus terbanyak selama 6 tahun terakhir yaitu sebanyak 683 kasus.

Data kasus perdagangan orang menggambarkan banyaknya kasus yang terjadi pada periode waktu tertentu, sehingga data kasus perdagangan orang merupakan data cacahan. *Generalized Linear Models* (GLM) dapat digunakan dalam pemodelan data cacahan. GLM merupakan pemodelan yang dapat digunakan ketika data tidak berdistribusi normal [10]. Regresi Poisson merupakan salah satu teknik regresi dalam GLM yang sering diterapkan dalam analisis data berbentuk cacahan dengan syarat asumsi equidispersi harus dipenuhi. Equidispersi menyatakan bahwa rata-rata dan varians variabel respon memiliki nilai yang sama. Salah satu metode alternatif yang dapat diterapkan ketika asumsi equidispersi tidak terpenuhi adalah regresi Binomial Negatif [8].

Regresi Binomial Negatif dapat diterapkan jika data cacahan dianggap dilaporkan secara akurat. Oleh karena itu, pada pemodelan data *underreported counts* diperlukan pengembangan model untuk mengestimasi nilai rata-rata banyaknya kejadian yang sebenarnya seperti model yang dikembangkan oleh Mukhopadhyay (1997), yaitu model regresi Binomial Negatif untuk *underreported counts* [5].

Estimasi parameter model dapat dilakukan melalui pendekatan Bayesian dengan menyertakan distribusi dari suatu parameter model. Pendekatan Bayesian dapat memberikan kesimpulan dan perkiraan yang stabil bahkan dalam situasi dengan ukuran sampel yang sangat kecil. Oleh karena itu, menggunakan pendekatan Bayesian merupakan pilihan terbaik mengenai parameter populasi dalam memperbaharui kepercayaan seseorang setelah mendapatkan informasi baru [4].

Berdasarkan latar belakang tersebut, penelitian ini memiliki tujuan untuk memprediksi jumlah kasus perdagangan perempuan yang sebenarnya di Indonesia tahun 2023 menggunakan regresi Binomial Negatif untuk *underreported counts*.

2. HASIL DAN PEMBAHASAN

2.1 Pemeriksaan Asumsi Equidispersi

Pemeriksaan asumsi equidispersi bertujuan untuk melihat apakah rata-rata dan varians dari variabel respon sama atau berbeda. Jika rata-rata lebih besar dari varians, maka terjadi pelanggaran asumsi equidispersi yang disebut overdispersi. Pemeriksaan asumsi equidispersi dapat dilakukan dengan persamaan sebagai berikut [3]:

$$\phi = \frac{D}{db} \quad (2.1)$$

$$D = 2 \sum_{i=1}^n \left(y_i \ln \left(\frac{y_i}{\hat{\lambda}_i} \right) - (y_i - \hat{\lambda}_i) \right) \quad (2.2)$$

$$db = n - p \quad (2.3)$$

Keterangan:

ϕ = Nilai dispersi

D = Nilai *deviance*

db = Derajat bebas

y_i = Nilai variabel dependen pengamatan ke- i

$\hat{\lambda}_i$ = Taksiran nilai rata-rata banyak kasus ke- i pada model regresi

p = Banyak parameter

n = Banyak pengamatan

Hipotesis dalam pemeriksaan asumsi equidispersi yaitu:

$H_0 : \phi \leq 1$ (Tidak terjadi overdispersi pada model regresi)

$H_1 : \phi > 1$ (Terjadi overdispersi pada model regresi)

Tabel 2.1 Pemeriksaan Asumsi Equidisersi

D	db	ϕ
157,7412	30	5,2581

Berdasarkan Tabel 2.1 nilai dispersi (5,2581) > 1 sehingga H_1 diterima. Oleh karena itu, dapat ditarik kesimpulan bahwa asumsi equidisersi tidak terpenuhi atau terjadi overdispersi, sehingga model yang digunakan yaitu model regresi Binomial Negatif.

2.2 Regresi Binomial Negatif Untuk *Underreported Counts*

Regresi Binomial Negatif untuk *underreported counts* merupakan solusi untuk mengatasi masalah *underreported counts* dengan kasus overdispersi [11]. Regresi Binomial Negatif untuk *underreported counts* berasal dari distribusi campuran Poisson-Gamma dengan y_i^* merupakan banyaknya kejadian sebenarnya untuk setiap individu yang berdistribusi Poisson dengan rata-rata λ_i , di mana λ_i antar individu berdistribusi Gamma dengan parameter (θ, θ) .

Distribusi marginal dari banyaknya kejadian sebenarnya (y_i^*) berdistribusi Binomial Negatif dengan parameter (λ_i, α) di mana α merupakan parameter dispersi dengan $\alpha = \frac{1}{\theta}$ [9].

$$P(Y^* = y_i^*) = \frac{\Gamma(y_i^* + \frac{1}{\alpha})}{y_i^*! \Gamma(\frac{1}{\alpha})} \left(\frac{1}{1 + \alpha\lambda_i} \right)^{\frac{1}{\alpha}} \left(\frac{\alpha\lambda_i}{1 + \alpha\lambda_i} \right)^{y_i^*} \quad (2.4)$$

Distribusi marginal dari banyaknya kejadian yang dilaporkan (y_i) berdistribusi Binomial Negatif dengan parameter $(\pi\lambda_i, \alpha)$ di mana rata-rata $E(y_i) = \pi\lambda_i$ dan varians $Var(y_i) = \pi\lambda_i(1 + \alpha\pi\lambda_i)$ [9].

$$P(Y = y_i) = \frac{\Gamma(y_i + \frac{1}{\alpha})}{y_i! \Gamma(\frac{1}{\alpha})} \left(\frac{1}{1 + \alpha\pi\lambda_i} \right)^{\frac{1}{\alpha}} \left(\frac{\alpha\pi\lambda_i}{1 + \alpha\pi\lambda_i} \right)^{y_i} \quad (2.5)$$

Model regresi Binomial Negatif *underreported counts* yaitu [9]:

$$\lambda_i = \exp(x_i^T \beta) \quad (2.6)$$

Di mana $i = 1, 2, \dots, n$; x_i merupakan nilai variabel independen ke- i dan λ_i merupakan rata-rata dari y_i^* , sedangkan β merupakan parameter regresi yang akan diestimasi dengan pendekatan Bayesian.

2.3 Penaksiran Parameter Model *Underreported Counts*

Terdapat 2 jenis pendekatan dalam penaksiran parameter yaitu pendekatan statistika klasik dan pendekatan Bayesian [4]. Pada pendekatan Bayesian diperlukan distribusi *prior*, yang merupakan distribusi dari suatu parameter model yang diasumsikan sebagai variabel acak. Selain itu, distribusi *posterior* pada pendekatan Bayesian dibutuhkan untuk mengestimasi parameter model hasil taksiran Bayesian yang diperoleh dari rata-rata distribusi *posterior*. Teorema Bayes umumnya didefinisikan sebagai berikut [4]:

$$posterior \propto likelihood \times prior \quad (2.7)$$

Dengan melakukan penjabaran untuk fungsi *likelihood* dan fungsi distribusi *prior* untuk model regresi Binomial Negatif *underreported counts*, maka distribusi *posterior* untuk model regresi Binomial Negatif *underreported counts* adalah sebagai berikut:

$$P(y_i^*, \lambda_i, \beta, \pi, \theta | y_i) \propto P(y_i | y_i^*, \pi) \times P(y_i^* | \beta) \times P(\lambda_i | \theta) \times f(\beta) \times f(\theta) \times f(\pi) \quad (2.8)$$

$$\begin{aligned}
P(y_i^*, \lambda_i, \beta, \pi, \theta | y_i) & \propto \prod_{i=1}^n \frac{y_i^*!}{(y_i^* - y_i)! y_i!} \pi^{y_i} (1 - \pi)^{y_i^* - y_i} \times \frac{\exp(y_i^* x_i' \beta - \exp(x_i' \beta))}{y_i^*!} \\
& \times \frac{\theta^\theta}{\Gamma(\theta)} \lambda_i^{\theta-1} e^{-\theta \lambda_i} \times \frac{1}{(\sqrt{2\pi})^m |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\beta - \mu)' \Sigma^{-1}(\beta - \mu)\right) \\
& \times \frac{1}{\Gamma(c_0) C^{c_0}} \theta^{c_0-1} e^{-\frac{\theta}{C}} \times 1
\end{aligned} \tag{2.9}$$

Persamaan (2.9) merupakan distribusi *posterior* yang akan digunakan. Namun, distribusi yang diperoleh cukup kompleks sehingga sulit untuk menentukan distribusi *posterior* dalam melakukan penaksiran parameter. Oleh sebab itu, penaksiran parameter model dikerjakan dengan menerapkan metode *Markov Chain Monte Carlo* (MCMC).

2.4 Markov Chain Monte Carlo (MCMC)

Markov Chain Monte Carlo (MCMC) yaitu metode simulasi yang digunakan untuk menarik sampel dari distribusi *posterior* dan menghitung nilai besaran yang diamati [4]. MCMC didasarkan pada pembentukan rantai markov yang cepat mencapai konvergen (stasioner) pada distribusi *posterior*. Dengan menggunakan algoritma dalam pengaplikasiannya, MCMC menghasilkan data sampel sebuah parameter yang berdistribusi tertentu. Algoritma ini dioperasikan secara iteratif hingga membentuk suatu rantai markov, di mana setiap nilai pada setiap langkah bergantung pada langkah sebelumnya. Salah satu algoritma MCMC yang sering diterapkan adalah *Gibbs sampling*. *Gibbs sampling* merupakan kasus khusus dari algoritma *Metropolis-Hasting* dengan probabilitas penerimaan selalu bernilai satu dan dapat diterapkan ketika distribusi bersyarat untuk masing-masing parameter tersedia [5].

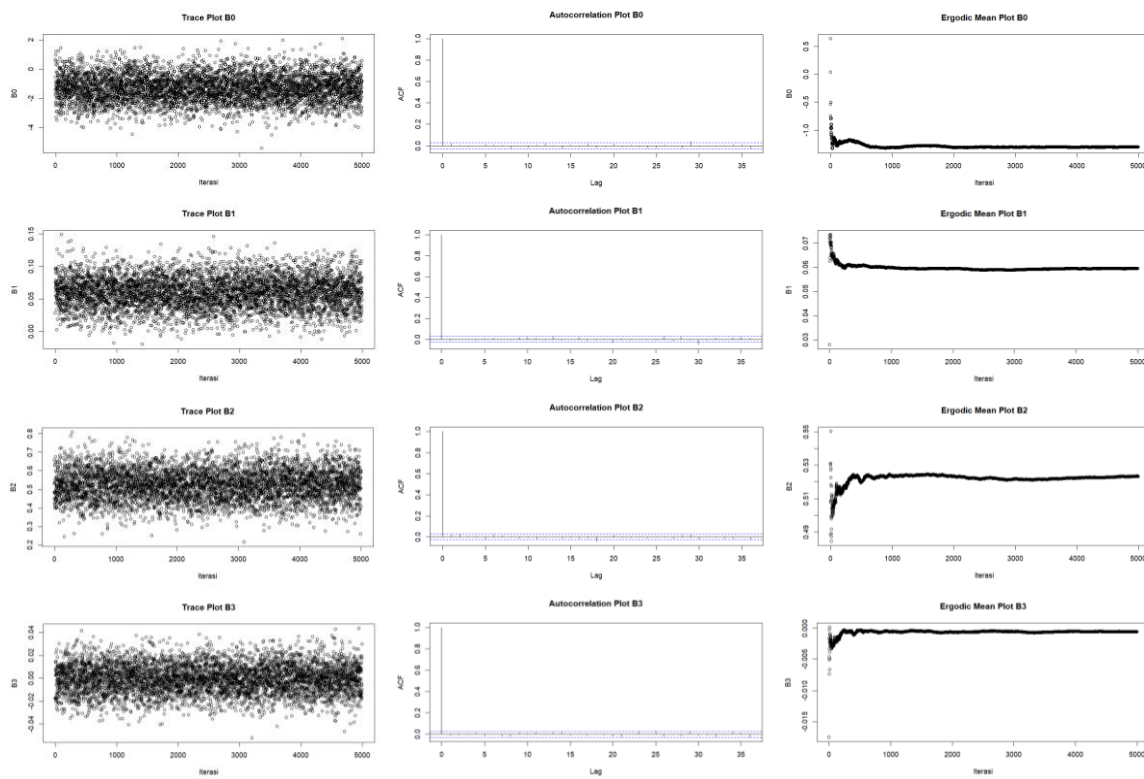
Estimasi parameter model regresi Binomial Negatif untuk *underreported counts* menggunakan MCMC dengan algoritma *Gibbs sampling* dilakukan dengan tahapan sebagai berikut:

1. Tentukan nilai awal untuk parameter model yaitu $\beta^{(0)}, \theta^{(0)}, \pi^{(0)}, y^*$
2. Tentukan banyaknya iterasi yang digunakan (T)
3. Jalankan simulasi untuk setiap iterasi $t = 1, 2, \dots, T$ dengan langkah-langkah sebagai berikut:
 - a. *Generate* kandidat nilai baru $y_i^{*(t)}$ dari fungsi distribusi bersyarat $y_i^* | \beta^{(t-1)}, \pi^{(t-1)}, y_i$
 - b. *Generate* kandidat nilai baru $\pi_i^{(t)}$ dari fungsi distribusi bersyarat $\pi_i | y_i, y_i^{*(t)}$
 - c. *Generate* kandidat baru parameter $\beta^{(t)}$ dari fungsi distribusi bersyarat $\beta | y_i^{*(t)}, y_i$ melalui algoritma *Random-Walks*
 - d. *Generate* kandidat baru parameter $\theta^{(t)}$ dari fungsi distribusi bersyarat $\theta | y_i, y_i^{*(t)}$ melalui algoritma *Random Walks*
4. Perbarui nilai-nilai parameter model berdasarkan nilai-nilai yang diperoleh dari hasil simulasi
5. Evaluasi konvergensi algoritma menggunakan *trace plot* (konvergensi tercapai jika semua nilai-nilai berada dalam sebuah daerah yang tidak menunjukkan periodisitas yang kuat dan kecenderungan membentuk sebuah pola), *autocorrelation plot* (konvergensi tercapai jika tidak terdapat korelasi yang kuat diantara iterasi yang beruntun), dan *ergodic mean plot* (algoritma dikatakan telah mencapai konvergen ketika *ergodic mean plot* menunjukkan pola yang stabil setelah beberapa iterasi).
6. Tetapkan *burn-in period*
7. Dari hasil simulasi mulai dari tahap *burn-in* hingga iterasi terakhir, hitung:
 - a. Nilai *mean* serta simpangan baku dari parameter β
 - b. Nilai *mean* serta simpangan baku dari parameter θ
 - c. Nilai *mean* serta simpangan baku dari parameter π_i untuk $i = 1, 2, \dots, n$

- d. Nilai *mean* serta simpangan baku dari parameter y_i^* untuk $i = 1, 2, \dots, n$

2.5 Pemeriksaan Konvergensi Algoritma

Pada metode MCMC diperlukan pemeriksaan konvergensi algoritma untuk mengetahui apakah algoritma yang digunakan telah mencapai kondisi konvergen. Pemeriksaan konvergensi algoritma dapat dilakukan melalui *trace plot*, *autocorrelation plot*, dan *ergodic mean plot* [4]. Pemeriksaan *trace plot* dilakukan untuk pendugaan *burn-in period* dengan melihat sebaran yang dihasilkan. Pemeriksaan *autocorrelation plot* dilakukan untuk melihat adanya korelasi di antara iterasi yang berurutan, di mana nilai korelasi yang rendah mengindikasikan bahwa algoritma telah berada di wilayah distribusi target. Pemeriksaan *ergodic mean plot* dilakukan untuk menentukan *burn-in period* dengan melihat apakah nilai *ergodic mean* sudah stabil setelah beberapa iterasi.

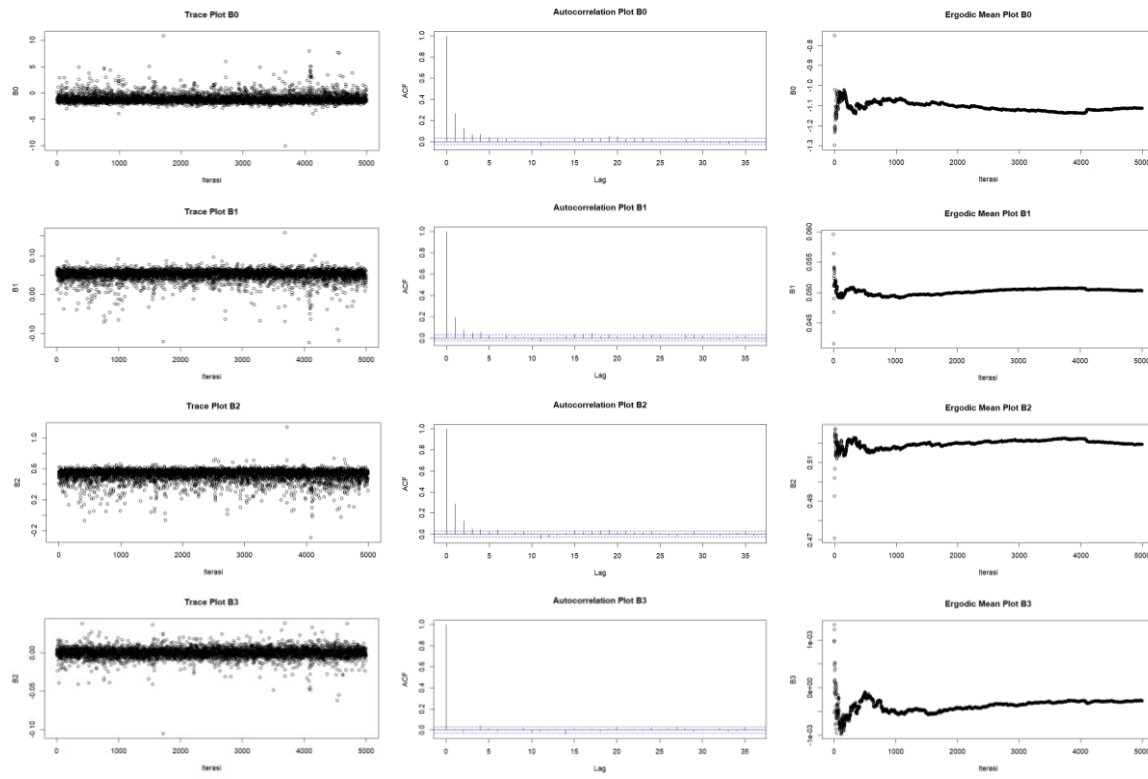


Gambar 2.1 Trace Plot, Autocorrelation Plot, Ergodic Mean Plot Random-Walks Parameter $\beta^{(t=1)}$

Gambar 2.1 merupakan *trace plot*, *autocorrelation plot*, dan *ergodic mean plot* iterasi pertama algoritma *Random-Walks* untuk parameter β . *Trace plot* menunjukkan bahwa plot tidak menunjukkan sebuah pola tertentu sehingga plot memiliki sebaran yang acak. *Autocorrelation plot* menunjukkan nilai autokorelasi pada *lag* pertama mendekati satu, dimana nilai-nilai korelasi selanjutnya semakin kecil dan menuju nol. *Ergodic mean plot* yang diperoleh memperlihatkan sebuah plot yang mulai stabil setelah 2000 iterasi pertama, sehingga dapat ditarik kesimpulan bahwa algoritma *Random-Walks* pada penaksiran parameter β telah mencapai konvergen dengan *burn-in period* pada 2000 iterasi. Oleh karena itu, taksiran parameter β dihitung dari iterasi ke-2001 sampai 5000 ($r = 2001, \dots, 5000$).

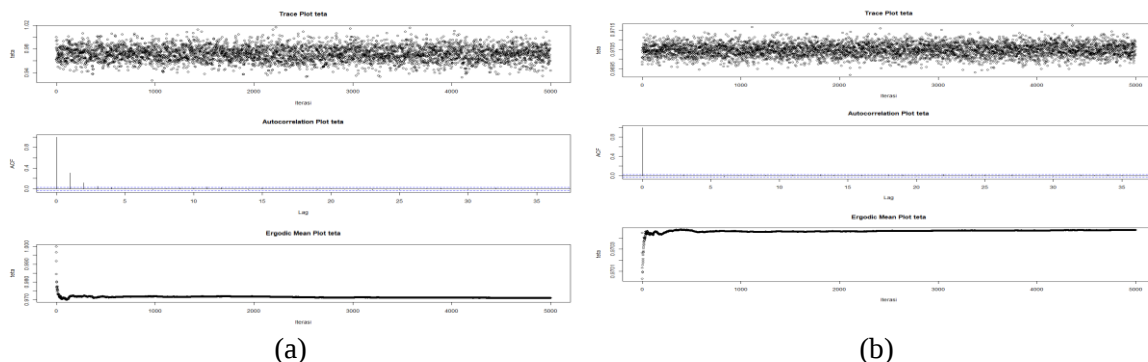
Penaksiran parameter dengan algoritma *Gibbs Sampling* dilakukan dengan simulasi *Random-Walks* parameter β untuk iterasi selanjutnya ($t = 2, \dots, 5000$) dengan *burn-in period* yang sama

yaitu 2000 iterasi. Dilakukan pemeriksaan konvergensi algoritma *gibbs sampling* melalui *trace plot*, *autocorrelation plot*, dan *ergodic mean plot* yang disajikan pada Gambar 2.2.



Gambar 2.2 Trace Plot, Autocorrelation Plot, Ergodic Mean Plot Random-Walks Parameter β

Gambar 2.2 merupakan *trace plot*, *autocorrelation plot*, dan *ergodic mean plot* algoritma *Random-Walks* untuk parameter β . *Trace plot* tidak memperlihatkan sebuah pola tertentu sehingga plot memiliki sebaran yang acak. *Autocorrelation plot* menunjukkan bahwa nilai autokorelasi pada lag pertama mendekati satu, sementara nilai-nilai pada lag berikutnya semakin menurun dan mendekati nol. *Ergodic mean plot* yang dihasilkan menggambarkan bahwa nilai *ergodic mean* membentuk pola yang mulai stabil setelah 3000 iterasi pertama, sehingga algoritma *Random-Walks* pada penaksiran parameter β telah konvergen serta mencapai *burn-in period* pada 3000 iterasi. Oleh karena itu, taksiran parameter β dihitung dari iterasi ke-3001 sampai 5000 ($t = 3001, \dots, 5000$).

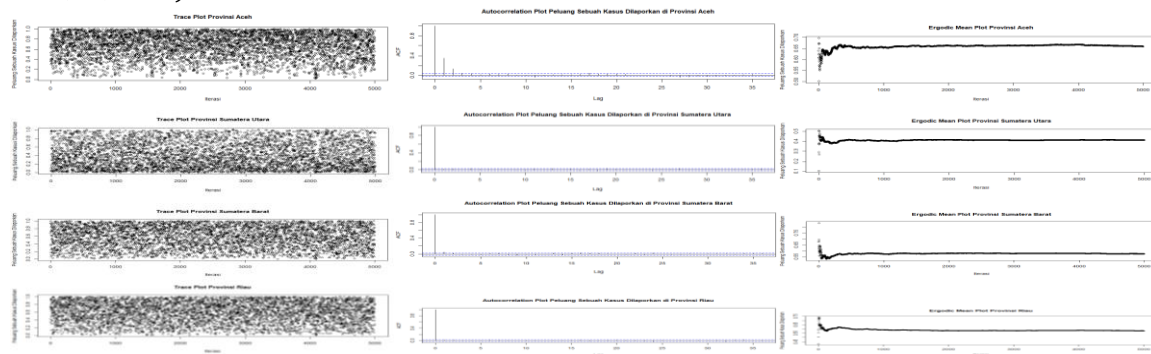


(a) (b)
Gambar 2.3 Trace Plot, Autocorrelation Plot, Ergodic Mean Plot Random-Walks
(a) Parameter $\theta^{(t=1)}$ dan (b) Parameter θ

Gambar 2.3(a) merupakan *trace plot*, *autocorrelation plot*, dan *ergodic mean plot* iterasi pertama algoritma *Random-Walks* untuk parameter θ . *Trace plot* tidak menggambarkan sebuah pola tertentu, artinya plot memiliki sebaran yang acak. *Autocorrelation plot* memperlihatkan bahwa nilai autokorelasi pada lag pertama hampir mencapai satu, dan nilai-nilai pada lag selanjutnya semakin kecil dan mendekati nol. *Ergodic mean plot* membentuk pola dari nilai *ergodic mean* yang mulai stabil setelah 2000 iterasi pertama. Oleh karena itu, algoritma *Random-Walks* pada penaksiran parameter θ mencapai konvergen dengan *burn-in period* pada 2000 iterasi. Artinya, taksiran parameter θ dihitung dari iterasi ke-2001 sampai 5000 ($h = 2001, \dots, 5000$).

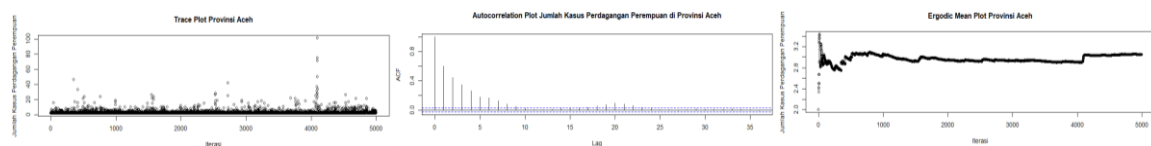
Penaksiran parameter dengan algoritma *Gibbs sampling* dilakukan dengan simulasi *Random-Walks* parameter θ untuk iterasi selanjutnya ($t = 2, \dots, 5000$) dengan *burn-in period* yang sama yaitu 2000 iterasi. Dilakukan pemeriksaan konvergensi algoritma *Gibbs Sampling* melalui *trace plot*, *autocorrelation plot*, dan *ergodic mean plot* yang disajikan pada gambar 2.3 poin (b).

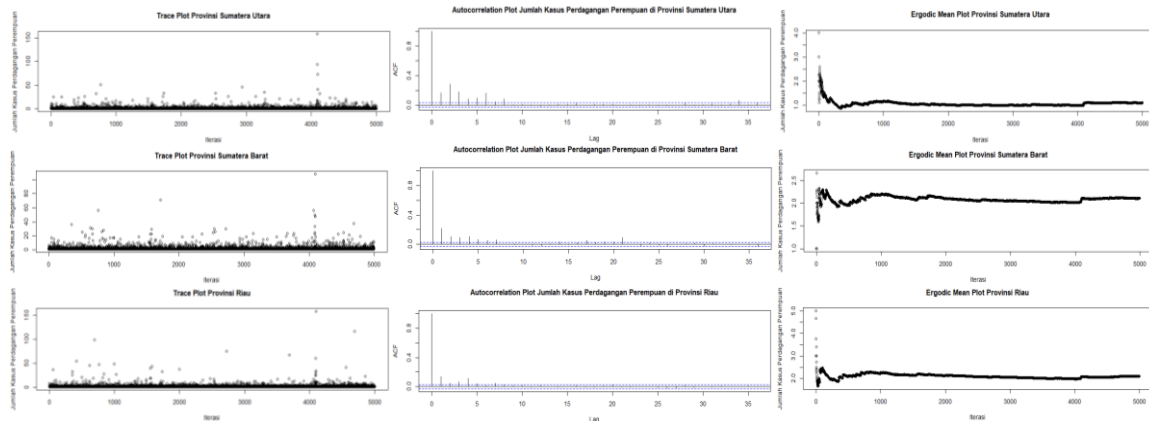
Gambar 2.3(b) merupakan *trace plot*, *autocorrelation plot*, dan *ergodic mean plot* algoritma *Random-Walks* untuk parameter θ . *Trace plot* menunjukkan bahwa plot tidak membentuk sebuah pola tertentu sehingga sebaran bersifat acak. *Autocorrelation plot* menunjukkan bahwa nilai autokorelasi pada lag pertama mendekati nilai satu, dan nilai-nilai pada lag berikutnya semakin kecil hingga mendekati nol. *Ergodic mean plot* mengindikasikan bahwa pola yang terbentuk dari nilai *ergodic mean* mulai stabil setelah 3000 iterasi pertama, sehingga dapat ditarik kesimpulan bahwa konvergensi algoritma *Random-Walks* pada penaksiran parameter θ mencapai *burn-in period* pada 3000 iterasi. Oleh karena itu, taksiran parameter θ dihitung dari iterasi ke-3001 sampai 5000 ($t = 3001, \dots, 5000$).



Gambar 2.4 Trace Plot, Autocorrelation Plot, Ergodic Mean Plot Random-Walks $\pi_1, \pi_2, \pi_3, \pi_4$

Gambar 2.4 merupakan *trace plot*, *autocorrelation plot*, dan *ergodic mean plot* algoritma *Random-Walks* untuk parameter π . *Trace plot* tidak memperlihatkan pola tertentu, sehingga distribusinya bersifat acak. *Autocorrelation plot* menunjukkan bahwa nilai autokorelasi pada lag pertama mendekati satu, sementara pada lag-lag berikutnya nilainya semakin kecil dan cenderung menuju nol. *Ergodic mean plot* yang terbentuk menunjukkan bahwa nilai *ergodic mean* sudah mulai stabil setelah 3000 iterasi pertama, sehingga dapat disimpulkan bahwa konvergensi algoritma *Random-Walks* pada penaksiran parameter π mencapai *burn-in period* pada 3000 iterasi. Oleh karena itu, taksiran parameter π dihitung dari iterasi ke-3001 sampai 5000 ($t = 3001, \dots, 5000$).





Gambar 2.5 Trace Plot, Autocorrelation Plot, Ergodic Mean Plot Random-Walks $y_1^*, y_2^*, y_3^*, y_4^*$

Gambar 2.5 merupakan *trace plot*, *autocorrelation plot*, dan *ergodic mean plot* algoritma *Random-Walks* untuk parameter y^* . *Trace plot* tidak menunjukkan sebuah pola tertentu yang mengindikasikan sebuah sebaran yang acak. *Autocorrelation plot* menunjukkan bahwa nilai autokorelasi semakin kecil dan mendekati nol. *Ergodic mean plot* menunjukkan bahwa nilai *ergodic mean* sudah mulai stabil setelah 3000 iterasi pertama, sehingga dapat disimpulkan bahwa konvergensi algoritma *Random-Walks* pada penaksiran parameter y^* mencapai *burn-in period* pada 3000 iterasi. Oleh karena itu, taksiran parameter y^* dihitung dari iterasi ke-3001 sampai 5000 ($t = 3001, \dots, 5000$).

2.6 Pemeriksaan Konvergensi Rantai

Dalam pendekatan Bayesian dengan menggunakan metode MCMC, diperlukan pengujian statistik untuk melihat kekonvergenan rantai markov. Salah satu statistik uji yang dapat digunakan untuk melakukan pemeriksaan konvergensi rantai adalah statistik Gelman-Rubin, sebagai berikut [2]:

$$\hat{R} = \sqrt{\frac{\hat{\sigma}_{\theta}^2}{W}} \quad (2.10)$$

$$\bar{\theta}_m = \frac{1}{N} \sum_{n=1}^N \theta_m^n \quad (2.11)$$

$$\bar{\theta} = \frac{1}{M} \sum_{m=1}^M \bar{\theta}_m \quad (2.12)$$

$$B = \frac{N}{M-1} \sum_{m=1}^M (\bar{\theta}_m - \bar{\theta})^2 \quad (2.13)$$

$$W = \frac{1}{M(N-1)} \sum_{m=1}^M \sum_{n=1}^N (\theta_m^n - \bar{\theta}_m)^2 \quad (2.14)$$

$$\hat{\sigma}_{\theta}^2 = \left(1 - \frac{1}{N}\right) W + \left(\frac{1}{N}\right) B \quad (2.15)$$

Keterangan:

N = Banyak iterasi yang dilakukan

M = Banyak pengulangan yang dilakukan

$\bar{\theta}_m$ = Rata-rata rantai markov ke- m

$\bar{\theta}$ = Rata-rata keseluruhan rantai markov

θ_m^n = Nilai parameter θ pada iterasi ke- n dari rantai markov ke- m

$\hat{\sigma}_\theta^2$ = Varians gabungan

Gelman menyebutkan bahwa jika statistik uji Gelman-Rubin yang diperoleh kurang dari 1,2 ($\hat{R} < 1,2$) maka cukup untuk menyimpulkan bahwa konvergensi antar rantai sudah tercapai [2].

Tabel 2.2 Statistik Gelman-Rubin Parameter β dan θ

Parameter	$\hat{R}$
β_0	1,0014
β_1	1,0014
β_2	1,0017
β_3	1,0003
θ	1,0917

Tabel 2.3 Statistik Gelman-Rubin Parameter π dan y^*

Parameter	$\hat{R}$	Parameter	$\hat{R}$	Parameter	$\hat{R}$	Parameter	$\hat{R}$
π_1	1,0011	π_{18}	1,0004	y_1^*	1,0025	y_{18}^*	1,0005
π_2	1,0003	π_{19}	1,0005	y_2^*	1,0019	y_{19}^*	1,0008
π_3	1,0022	π_{20}	1,0004	y_3^*	1,0011	y_{20}^*	1,0012
π_4	1,0004	π_{21}	1,0011	y_4^*	1,0012	y_{21}^*	1,0018
π_5	1,0009	π_{22}	1,0002	y_5^*	1,0017	y_{22}^*	1,0012
π_6	1,0014	π_{23}	1,0005	y_6^*	1,0031	y_{23}^*	1,0007
π_7	1,0009	π_{24}	1,0008	y_7^*	1,0027	y_{24}^*	1,0009
π_8	1,0008	π_{25}	1,0002	y_8^*	1,0014	y_{25}^*	1,0011
π_9	1,0009	π_{26}	1,0003	y_9^*	1,0013	y_{26}^*	1,0025
π_{10}	1,0003	π_{27}	1,0020	y_{10}^*	1,0005	y_{27}^*	1,0018
π_{11}	1,0001	π_{28}	1,0002	y_{11}^*	1,0017	y_{28}^*	1,0003
π_{12}	1,0008	π_{29}	1,0020	y_{12}^*	1,0006	y_{29}^*	1,0025
π_{13}	1,0009	π_{30}	1,0008	y_{13}^*	1,0016	y_{30}^*	1,0026
π_{14}	1,0004	π_{31}	1,0005	y_{14}^*	1,0023	y_{31}^*	1,0012
π_{15}	1,0019	π_{32}	1,0009	y_{15}^*	1,0014	y_{32}^*	1,0014
π_{16}	1,0011	π_{33}	1,0008	y_{16}^*	1,0014	y_{33}^*	1,0015
π_{17}	1,0007	π_{34}	1,0001	y_{17}^*	1,0009	y_{34}^*	1,0010

Berdasarkan Tabel 2.2 dan 2.3, diperoleh statistik Gelman-Rubin parameter β, θ, π, y^* lebih dari 1 dan kurang dari 1,2 ($1 < \beta, \theta, \pi, y^* < 1.2$) maka cukup untuk menyimpulkan bahwa konvergensi antar rantai parameter β, θ, π, y^* sudah tercapai. Oleh karena itu, taksiran parameter β, θ, π, y^* dihitung dengan menggabungkan nilai simulasi hasil iterasi parameter β, θ, π, y^* pada 5 rantai markov kemudian dihitung nilai rata-ratanya.

2.7 Penaksiran Parameter dan Pemodelan Regresi Binomial Negatif *Underreported Counts*

Penaksiran parameter dilakukan ketika konvergensi dari algoritma yang digunakan sudah dicapai.

JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI

Nabilah Disya Arviani, Reny Rian Marlina

Nilai taksiran parameter merupakan rata-rata dari nilai-nilai sampel hasil simulasi, di mana rumus taksiran parameter γ adalah sebagai berikut[4]:

$$\hat{\gamma}_i = \frac{\sum_{t=1}^{T'} \gamma_i^{(t)}}{T'} \quad (2.16)$$

Di mana $T' = T - B$ dan $i = 1, 2, \dots, k$

$\hat{\gamma}_i$ = Nilai taksiran parameter γ

$\gamma_i^{(t)}$ = Nilai sampel hasil simulasi

T = Banyak iterasi yang dilakukan

B = Banyak iterasi yang dihilangkan pada *burn-in period*

Tabel 2.4 Taksiran Parameter Regresi Binomial Negatif *Underreported Counts*

Parameter	Taksiran	Standar Deviasi
β_0	-1,1263	0,3655
β_1	0,0503	0,0065
β_2	0,5193	0,0359
β_3	-0,0001	0,0029

Berdasarkan Tabel 2.4, model regresi Binomial Negatif *underreported counts* menunjukkan taksiran rata-rata jumlah kasus perdagangan perempuan yang sebenarnya di Indonesia tahun 2023 untuk provinsi ke- i di mana $i = 1, 2, \dots, 34$, sebagai berikut:

$$\hat{\lambda}_i = \exp(-1,1263 + 0,0503X_{1i} + 0,5193X_{2i} - 0,0001X_{3i})$$

Berdasarkan model regresi tersebut, $\hat{\lambda}_i$ menunjukkan taksiran rata-rata jumlah kasus perdagangan perempuan pada provinsi ke- i yang memiliki satuan jiwa yang tidak mungkin bernilai koma, sehingga dilakukan pembulatan ke atas pada nilai taksiran rata-rata. Berikut merupakan taksiran rata-rata jumlah kasus perdagangan perempuan yang sebenarnya di Indonesia tahun 2023:

Tabel 2.5 Taksiran Rata-Rata Jumlah Kasus Perdagangan Perempuan yang Sebenarnya di Indonesia Tahun 2023

Provinsi	Taksiran Rata-Rata	Provinsi	Taksiran Rata-Rata
Aceh	16	Nusa Tenggara Barat	3
Sumatera Utara	11	Nusa Tenggara Timur	5
Sumatera Barat	10	Kalimantan Barat	7
Riau	5	Kalimantan Tengah	4
Jambi	5	Kalimantan Selatan	4
Sumatera Selatan	5	Kalimantan Timur	7
Bengkulu	4	Kalimantan Utara	4
Lampung	6	Sulawesi Utara	12
Bangka Belitung	5	Sulawesi Tengah	3
Kepulauan Riau	15	Sulawesi Selatan	5
DKI Jakarta	12	Sulawesi Tenggara	3
Jawa Barat	23	Gorontalo	4
Jawa Tengah	8	Sulawesi Barat	2
DI Yogyakarta	4	Maluku	20
Jawa Timur	7	Maluku Utara	5
Banten	22	Papua	5
Bali	2	Papua Barat	15

JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI

Nabilah Disya Arviani, Reny Rian Marlina

Berdasarkan Tabel 2.5 dapat dilihat bahwa Provinsi Jawa Barat memiliki rata-rata jumlah kasus perdagangan perempuan yang sebenarnya paling tinggi pada tahun 2023 yaitu sebanyak 23 kasus, sedangkan rata-rata jumlah kasus perdagangan perempuan yang sebenarnya paling rendah pada tahun 2023 berada di Provinsi Bali dan Sulawesi Barat yaitu sebanyak 2 kasus.

Tabel 2.6 Taksiran Parameter Dispersi

Parameter	Taksiran
θ	0,9704
α	1,0305

Berdasarkan Tabel 2.6 taksiran parameter dispersi (α) sebesar 1,0305 menunjukkan bahwa data memiliki variabilitas tambahan yang tidak bisa dijelaskan hanya oleh rata-rata sekitar 3,05% dibandingkan dengan model regresi Poisson.

Tabel 2.7 Taksiran Peluang Sebuah Kasus Perdagangan Perempuan Dilaporkan

Provinsi	$\hat{\pi}_i$	$std(\hat{\pi}_i)$	$(1 - \hat{\pi}_i) \times 100\%$
Aceh	0,6614	0,1063	33,86%
Sumatera Utara	0,4190	0,1320	58,10%
Sumatera Barat	0,5795	0,1207	42,05%
Riau	0,5792	0,1190	42,08%
Jambi	0,8415	0,0668	15,85%
Sumatera Selatan	0,7567	0,0876	24,33%
Bengkulu	0,5759	0,1172	42,41%
Lampung	0,7870	0,0820	21,30%
Bangka Belitung	0,5761	0,1195	42,39%
Kepulauan Riau	0,9368	0,0329	6,32%
DKI Jakarta	0,4160	0,1279	58,40%
Jawa Barat	0,9582	0,0237	4,18%
Jawa Tengah	0,8428	0,0645	15,72%
DI Yogyakarta	0,7212	0,0973	27,88%
Jawa Timur	0,9053	0,0447	9,47%
Banten	0,8253	0,0697	17,47%
Bali	0,4212	0,1317	57,88%
Nusa Tenggara Barat	0,8640	0,0586	13,60%
Nusa Tenggara Timur	0,7572	0,0873	24,28%
Kalimantan Barat	0,6625	0,1039	33,75%
Kalimantan Tengah	0,5760	0,1194	42,40%
Kalimantan Selatan	0,5742	0,1168	42,58%
Kalimantan Timur	0,7142	0,0978	28,58%
Kalimantan Utara	0,8429	0,0646	15,71%
Sulawesi Utara	0,8525	0,0629	14,75%
Sulawesi Tengah	0,6648	0,1077	33,52%
Sulawesi Selatan	0,9088	0,0443	9,12%
Sulawesi Tenggara	0,4180	0,1291	58,20%
Gorontalo	0,7585	0,0884	24,15%
Sulawesi Barat	0,5778	0,1210	42,22%
Maluku	0,9585	0,0243	4,15%
Maluku Utara	0,7155	0,0957	28,45%
Papua	0,4153	0,1307	58,47%

JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI

Nabilah Disya Arviani, Reny Rian Marlina

Papua Barat	0,4197	0,1303	58,03%
Rata-Rata	0,6907	0,0919	30,93%

Berdasarkan taksiran peluang sebuah kasus perdagangan perempuan dilaporkan yang diperoleh pada Tabel 2.7 dapat dilihat bahwa Provinsi Maluku memiliki peluang sebuah kasus perdagangan perempuan dilaporkan paling tinggi pada tahun 2023 yaitu sebesar 0,9585 dengan angka persentase sebuah kasus perdagangan perempuan tidak dilaporkan paling kecil yaitu sebesar 4,15%. Sedangkan peluang sebuah kasus perdagangan perempuan dilaporkan paling rendah pada tahun 2023 berada di Provinsi Papua yaitu sebesar 0,4153 dengan angka persentase sebuah kasus perdagangan perempuan tidak dilaporkan paling besar yaitu sebesar 58,47%.

Tabel 2.8 Taksiran Jumlah Kasus Perdagangan Perempuan yang Sebenarnya di Indonesia Tahun 2023

Provinsi	y_i	$\hat{y}_i^*$	$std(\hat{y}_i^*)$	$\hat{y}_i^* - y_i$
Aceh	2	4	2	2
Sumatera Utara	0	2	2	2
Sumatera Barat	1	3	2	2
Riau	1	3	2	2
Jambi	8	10	2	2
Sumatera Selatan	4	6	2	2
Bengkulu	1	2	2	1
Lampung	5	6	2	1
Bangka Belitung	1	3	2	2
Kepulauan Riau	26	28	2	2
DKI Jakarta	0	2	3	2
Jawa Barat	42	44	2	2
Jawa Tengah	8	10	2	2
DI Yogyakarta	3	5	2	2
Jawa Timur	16	18	2	2
Banten	7	9	2	2
Bali	0	2	2	2
Nusa Tenggara Barat	10	12	2	2
Nusa Tenggara Timur	4	6	2	2
Kalimantan Barat	2	4	2	2
Kalimantan Tengah	1	3	2	2
Kalimantan Selatan	1	3	2	2
Kalimantan Timur	3	5	2	2
Kalimantan Utara	8	9	2	1
Sulawesi Utara	9	11	2	2
Sulawesi Tengah	2	4	3	2
Sulawesi Selatan	17	19	2	2
Sulawesi Tenggara	0	2	2	2
Gorontalo	4	6	2	2
Sulawesi Barat	1	3	3	2
Maluku	42	44	2	2
Maluku Utara	3	5	2	2
Papua	0	2	2	2
Papua Barat	0	1	2	1
Rata-Rata	7	9	3	2

Berdasarkan Tabel 2.8 dapat dilihat bahwa Provinsi Maluku memiliki taksiran jumlah kasus perdagangan perempuan yang sebenarnya paling tinggi pada tahun 2023 yaitu sebanyak 44 kasus, sedangkan taksiran jumlah kasus perdagangan perempuan yang sebenarnya paling rendah pada tahun 2023 berada di Provinsi Papua Barat yaitu sebanyak 1 kasus. Taksiran jumlah kasus perdagangan perempuan yang tidak dilaporkan di Indonesia pada tahun 2023 berkisar antara 1 dan 2 kasus.

4. KESIMPULAN

Berdasarkan analisis yang telah dilakukan, dapat disimpulkan bahwa kasus perdagangan perempuan merupakan data yang mengalami *underreported counts*. Oleh karena itu, dilakukan penaksiran jumlah kasus perdagangan perempuan yang sebenarnya di Indonesia tahun 2023 berdasarkan persentase penduduk miskin, tingkat pengangguran terbuka, dan tingkat penyelesaian pendidikan jenjang SMA/ sederajat menggunakan regresi Binomial Negatif *underreported counts* di mana penaksiran parameter dilakukan melalui pendekatan Bayesian dengan metode MCMC dan algoritma *Gibbs Sampling*.

Penaksiran parameter dilakukan dengan 5000 iterasi, di mana *trace plot*, *autocorrelation plot*, dan *ergodic mean plot* pada masing-masing parameter menunjukkan bahwa algoritma *gibbs sampling* mencapai *burn-in period* setelah 3000 iterasi. Berdasarkan taksiran jumlah kasus perdagangan perempuan yang sebenarnya di Indonesia tahun 2023, diperoleh hasil bahwa Provinsi Papua Barat memiliki jumlah kasus perdagangan perempuan yang sebenarnya paling tinggi pada tahun 2023 yaitu sebanyak 44 kasus, sedangkan jumlah kasus perdagangan perempuan yang sebenarnya paling rendah pada tahun 2023 berada di Provinsi Papua Barat yaitu sebanyak 1 kasus.

Rata-rata jumlah kasus perdagangan perempuan yang sebenarnya di Indonesia tahun 2023 yaitu sebanyak 9 kasus. Kasus perdagangan perempuan yang tidak dilaporkan merupakan selisih antara jumlah kasus yang sebenarnya dengan jumlah kasus yang dilaporkan, di mana rata-rata jumlah kasus perdagangan perempuan yang tidak dilaporkan di Indonesia tahun 2023 yaitu sebanyak 2 kasus.

DAFTAR PUSTAKA

- [1] Abdullah, R. H., 2019. Tinjauan Viktimologis terhadap Tindak Pidana Perdagangan Orang (Human Trafficking). *Jurnal Yustika: Media Hukum dan Keadilan*, Vol. 22, No. 01, 55-63. <https://doi.org/10.24123/yustika.v22i01.1958>.
- [2] Gamerman, D., & Lopes, H. F., 2006. *Markov chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, Second Edition. Chapman and Hall/CRC. <https://doi.org/10.1201/9781482296426>.
- [3] Harahap, R. J., Herrhyanto, N., & Priatna, B. A., 2019. Penerapan Data Count dengan Menggunakan Regresi Hurdle Poisson. *Jurnal EurekaMatika*, Vol. 7, No. 1, 11-23. <https://doi.org/10.17509/jem.v7i1.17884>.
- [4] Johannes, H. V., Padmadisastra, S., & Tantular, B., 2017. Meningkatkan Ketahanan Wilayah Melalui Estimasi Underreported Data Kejahatan Dengan Pendekatan Bayes. *Jurnal Ketahanan Sosial*, Vol. 23, No. 3, 376-394. <https://doi.org/10.22146/jkn.29197>.
- [5] Marlina, R. R., & Padmadisastra, S., 2018. Poisson Regression of Damage Product Sales Using MCMC. *Indonesian Journal of Statistics and Its Applications*, Vol. 2, No. 1, 1-12. <https://doi.org/10.29244/ijsa.v2i1.53>.
- [6] Pararai, M., Famoye, F., & Lee, C., 2010. Generalized Poisson-Poisson Mixture Model for Misreported Counts with an Application to Smoking Data. *Journal of Data Science*, Vol. 8, No. 4, 607-617. [https://doi.org/10.6339/JDS.2010.08\(4\).608](https://doi.org/10.6339/JDS.2010.08(4).608).
- [7] Plantika, Y., 2019. Faktor Penyebab Perdagangan Orang di Wilayah Hukum Polres Malang Kota. *Dialektika*, Vol. 14, No. 1, 9-15. <https://doi.org/10.20473/jsd.v14i1.2019.9-15>.

- [8] Prahutama, A., Sudarno, S., Suparti, S., & Mukid, M. A., 2017. Analisis Faktor-Faktor yang Mempengaruhi Angka Kematian Bayi di Jawa Tengah Menggunakan Regresi Generalized Poisson dan Binomial Negatif. *Statistika*, Vol. 5, No. 2, 1-6. <https://doi.org/10.26714/jsunimus.5.2.2017.%25p>.
- [9] Rahimighazikalayeh, G., 2018. *Adjusting For Mis-Reporting In Count Data* (Doctoral dissertation, University of South Carolina).
- [10] Sofyan, W., 2020. Pemodelan Angka Kematian Bayi di Provinsi Jawa Barat Menggunakan Metode Regresi Poisson Inverse Gaussian (PIG). Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Muhammadiyah Semarang.
- [11] Winkelmann, R., 2008. *Econometric Analysis of Count Data*. Springer Science & Business Media. <https://doi.org/10.1007/978-3-540-78389-3>.