ESTIMASI PARAMETER DAN PENGUJIAN HIPOTESIS PADA MODEL REGRESI GAMMA (STUDI KASUS: PEMODELAN PENCEMARAN SUNGAI DI SURABAYA)

Alvi Sahrin Nasution[1,a], Purhadi[2,b] and Sutikno[2,c]

[1]*Univeristas Graha Nusantara, Padangsdidimpuan, Indonesia.*

[2] *Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam*
Institut Sepuluh November, Kampus ITS Sukolilo, Surabaya, 60111, Indonesia.

e-mail:[a)]alvi14@mhs.statistika.its.ac.id; [b)]purhadi@statistika.its.ac.id ;[c)]sutikno@statistika.its.ac.id

Abstract

Analysy is regression constitutes beneficent statistical methods to model relationship among variable response gets prediktor Gamma and variable distribution. Regression model in a general way is built bases Normal assumption, but empirik ala frequent assumption to be breached. While patterns asymmetry data, analisis is classic especially with men inferensi statistic to model parameter doesn't give to usufruct the better, since this distribution is designed as distribution that flexible and adaptif. More approaching thus efficient and not require data normalization can be done by use of other distribution which is Gamma distribution. This research determines esti masi parameter and hypthosts testing of Gamma regression model utilizes Maximum Likelihood Estimation (MLE) and Weighted Least Square (WLS). Estimator that acquired on Gamma regression model this is vector g gradient with variable k. Since acquired result not close form, therefore to determine estimatornya shall utilize iteration method. Iteration method that is utilized is Raphson Sehinggamemerlukan's Newton iteration method first derivative and second generation to parameter to form Hessian's matrix. While in hypthosts testing that provided by Likelihood Ratio Test (LRT) utilized is concomitant quiz and partial quiz with distribution statistic quiz Chisquare. For parameter estimation by methodics MLE and WLS IS variable that signifikan is river water speed sehing ga variable the most regard river sacrilege at Surabaya with indicator *Biochemical Oxygen Demand* (BOD) are river speed

Key word : *Biochemical Oxygen Demand, Gamma regression, Maximum Like li-*

Abstrak

Analisis regresi merupakan metode statistik yang berguna untuk memodelkan hubungan antara variabel respon berdistribusi Gamma dan variabel prediktor. Model regresi pada umumnya dibangun berdasarkan asumsi Normal, tapi secara empirik asumsi sering terlanggar. Ketika pola data asimetri, analisis klasik terutama dengan menggunakan statistik inferensi terhadap parameter model tidak memberikan hasil yang lebih baik, karena distribusi ini didesain sebagai distribusi yang fleksibel dan adaptif. Dengan demikian pendekatan yang lebih efisien dan tidak memerlukan penormalan data dapat dilakukan dengan menggunakan distribusi lain yaitu distribusi Gamma. Penelitian ini menentukan estimasi parameter dan pengujian hipotesis dari model regresi Gamma menggunakan Maximum Likelihood Estimation (MLE) dan Weighted Least Square (WLS). Estimator yang diperoleh pada model regresi Gamma ini adalah vektor g gradien dengan variabel k. Karena hasil yang diperoleh tidak close form, maka untuk menentukan estimatornya harus menggunakan metode iterasi. Metode iterasi yang digunakan adalah metode iterasi Newton-Raphson sehinggamemerlukan turunan pertama dan turunan kedua terhadap parameter untuk membentuk matriks Hessian. Sementara dalam pengujian hipotesis yang disediakan oleh Likelihood Ratio Test (LRT) digunakan adalah uji serentak dan uji parsial dengan uji statistik distribusi Chi-square. Untuk estimasi parameter dengan metode MLE dan WLS variabel yang signifikan adalah kecepatan air sungai sehingga variabel yang paling mempengaruhi pencemaran sungai di Surabaya dengan indikator *Biochemical*

hood Estimation, Weighted Least Square, Raphson's newton.

Oxygen Demand (BOD) adalah kecepatan sungai

Kata kunci : *Biochemical Oxygen Demand, Regresi Gamma, Maximum Like lihood Estimation, Weighted Least Square, Newton Raphson.*

PENDAHULUAN

Fungsi gamma pertama kali diperkenalkan oleh ahli matematika berkebangsaaan swiss Leonhard Euler (1707-1783) yang bertujuan untuk mengeneralisasi faktorial pada bilangan non bulat. Karena fungsi ini dianggap penting sehingga banyak ahli matematika lainnya yang mempelajarinya dan mengembangkannya. Fungsi gamma akan membentuk distribusi gamma [1].Distribusi gamma sering dipakai pada model probabilitas untuk waktu tunggu, distribusi ini aplikasinya ditemukan pada beberapa contoh kasus dalam kehidupan sehari-hari misalkan pemodelan curah hujan pada dua alat pengukur pada lokasi tertentu [2], ketergantungan antara debit sungai tahunan dan curah hujan pada suatu lokasi [3], ketergantungan cu-rah hujan dengan kemarau [4]. Distribusi Gamma memiliki parameter bentuk, parameter skala dan parameter lokasi, dimana disimbolkan sdengan α, θ, dan γ.

Model regresi pada umumnya dibangun berdasarkan asumsi mengikuti distribusi Normal, namun pada praktiknya secara empirik, asumsi ini tidak selalu tepat karena mungkin saja distribusi data bersifat asimetris dan bahkan bisa juga lebih tebal atau berekor lebih tipis dari distribusi normal. Ada beberapa distribusi data yang relaksasinya mampu menangkap pola asimetris dan ketebalan pada ekor salah satu datanya adalah distribusi Gamma. Analisis klasik terutama dengan inferensi statistiknya terhadap parameter model tidak akan memberikan hasil yang lebih baik, oleh sebab itu distribusi Gamma dirancang untuk mengatasi pola data yang asimetri karena distribusi ini didesain sebagai distribusi yang fleksibel dan adaptif,

dengan demikian pendekatan yang lebih efisien dan tidak memerlukan penormalan data dapat diperoleh [5].

Penelitian lain yang berkaitan dengan penelitian ini diantaranya meneliti tentang model regresi Gamma dengan menggunakan dua metode penaksiran yaitu *Maximum Likelihood Estimationdan Weighted Least Square* [6]. *Generalized gamma distribution* (GGD) yaitu dalam kasus pengembangan pengklasifikasian tekstur gambar dan memodelkan koefisien dari panjang gelombang yang diperoleh dari algoritma untuk menghitung centroid dari beberapa parameter [7].

Salah tujuan dari penelitian ini yaitu menentukan estimator parameter dan melakukan pengujian terhadap model regresi Gamma. Kemudian mengaplikasikan regresi Gamma untuk mengetahui faktor-faktor yang mempengaruhi *Biochemical Oxigen Demand* (BOD) sungai di Surabaya. Variabel respon yaitu kadar BOD dalam penelitian ini mengikuti distribusi Gamma dan variabel prediktor merupakan karakteristik sungai yang terdiri dari lebar air sungai, kedalaman air sungai, kecepatan air sungai dan debit air sungai.

TINJAUAN PUSTAKA

A. Distribusi Gamma

Distribusi Gamma sering dipakai pada model probabilitas untuk waktu tunggu. Distribusi Gamma terbagi menjadi satu, dua dan tiga parameter [8]. Fungsi kepadatan probabilitas (FKP) dari satu parameter adalah

$$f(y|\alpha)=\frac{1}{\Gamma(\alpha)}y^{\alpha-1}e^{-y};\alpha>0;\ 0<y<\infty \quad (1)$$

Fungsi kepadatan probabilitas dari dua parameter adalah

$$f(y|\alpha,\theta)=\frac{1}{\Gamma(\alpha)\theta^{\alpha}}\mathrm{y}^{\alpha-1}\mathrm{e}^{-\frac{y}{\theta}}\ ;\ \alpha,\theta>0;\ 0<\mathrm{y}<\infty \quad (2)$$

Fungsi kepadatan probabilitas tiga parameter adalah

$$f(y|\alpha,\theta,\gamma)=\frac{1}{\Gamma(\alpha)\theta^{\alpha}}(\mathrm{y}-\gamma)^{\alpha-1}\mathrm{e}^{-(y-\gamma)/\theta}\ ;$$
$$\alpha,\theta>0;\ \gamma<\mathrm{y};\ \gamma<\mathrm{y}<\infty \quad (3)$$

Dimana α merupakan parameter bentuk, θ merupakan parameter skala dan γ merupakan parameter lokasi. Terlepas dari adanya distribusi Gamma dengan tiga parameter, pada umumnya yang digunakan adalah distribusi Gamma dua parameter.

B. Regresi Gamma

Regresi Gamma adalah salah satu regresi yang dapat menggambarkan hubungan antara variabel y sebagai variabel respon berdistribusi Gamma dengan variabel x_1, x_2, … , x_k. Univariat dapat diartikan bahwa jumlah variabel respon hanya satu. Bentuk matriks dari variabel respon, variabel prediktor dan parameter regresi gamma adalah

$$\mathbf{y}=[y_1,y_2,\cdots,y_k]^T{}_{(\mathrm{nx1})};\mathbf{x}=[x_1,x_2,\cdots,x_k]^T{}_{(\mathrm{nx1})}$$
$$\boldsymbol{\beta}=[\beta_0,\beta_1,\cdots,\beta_k]^T{}_{(\mathrm{nx1})}$$

Maka fungsi kepadatan probabilitas dari model regresi Gamma adalah

$$f_{y_i}(y_i)=\frac{1}{\Gamma(\alpha)\left(\frac{\exp(\mathbf{x}_i{}^{\mathbf{T}}\boldsymbol{\beta})}{\alpha}\right)^{\alpha}}y_i^{\alpha-1}$$
$$\exp\left(-y_i\Big/\left(\frac{\exp(\mathbf{x}_i{}^{\mathbf{T}}\boldsymbol{\beta})}{\alpha}\right)\right) \quad (4)$$

Pada penelitian ini metode estimasi yang digunakan adalah MLE dan WLS. Estimasi parameter model regresi Gamma menggunakan metode estimasi maksimum likelihood (MLE) pada model regresi Gamma yaitu dengan memaksimumkan fungsi likelihood sehingga diperoleh fungsi likelihoodnya [9].

$$L(\boldsymbol{\beta}|\alpha)=\left(\frac{1}{\Gamma(\alpha)}\right)^{n}\left(\frac{\exp\sum_{i=1}^{n}\mathbf{x}_i{}^{\mathbf{T}}\boldsymbol{\beta}}{\alpha^{n}}\right)^{-\alpha}\left(\prod_{i=1}^{n}y_i{}^{\alpha-1}\right)$$
$$\exp\sum_{i=1}^{n}\left(-y_i\Big/\left(\frac{\exp(\mathbf{x}_i{}^{\mathbf{T}}\boldsymbol{\beta})}{\alpha}\right)\right) \quad (5)$$

Estimasi Weighted Least Square (WLS) adalah mengestimasi suatu garis regresi dengan jalan meminimalkan jumlah dari kuadrat kesalahan setiap observasi terhadap garis tersebut dengan cara membagi persamaan regresi OLS 10].

$$\hat{\beta}_w=\left(X^TD_wX\right)^{-1}X^TD_wZ \quad (6)$$

C. Pengujian Distribusi

Pengujian distribusi pada variabel respon menggunakan uji Anderson Darling merupakan salah satu metode statistik yang digunakan dalam pengujian kesesuaian distribusi. Uji Anderson Darling digunakan sebagai uji kenormalan atau goodness of fit untuk variabel kuantitatif dan bisa digunakan untuk menguji kenormalan berbagai macam sebaran data lognormal, normal, weibul, gamma dan sebaran logistik.

Adapun Hipotesis yang digunakan yaitu:

H_0: F(y) = variabel dependen sesuai dengan distribusi dugaan.

H_1: F(y) = variabel dependen tidak sesuai Dengan distribusi dugaan.

Statistik uji : (Shin, Jung, Jeong, & Heo, 2011)

$$A_n^2=-n-\frac{1}{n}\sum_{i=1}^{n}(2i-1)(\ln F^0_{(y_i)}+\ln(1-F^0_{(y+1-i)})) \quad (7)$$

$F^0_{(yi)}$ = Fungsi distributif kumulatif dari distribusi dugaan

n = ukuran sampel

Kriteria penolakan : Tolak H_0 jika nilai $A^2_{hit} > A^2_{\alpha}$

D. Pengujian Multikolineritas

Salah satu syarat harus dipenuhi dalam regresi dengan beberapa variabel prediktor adalah tidak adanya korelasi antara satu variabel prediktor dengan variabel prediktor lainnya. Multikolineritas adalah kondisi terdapatnya hubungan linear atau korelasi yang tinggi antara masing-masing variabel prediktor dalam model regresi [11].

Korelasi adalah tingkat tingginya keeretan hubungan dua variabel atau lebih yang digambarkan oleh besarnya koefisien korelasi. Koefisien korelasi adalah koefisien yang menggambarkan tingkat keeratan hubungan antar dua variabel atau lebih tetapi menggambarkan keterkaitan linear antar variabel [12]. Selain itu adanya kasus multikolineritas dapat dilihat menurut Variance Inflation Factor (VIF) [13].

$$VIF = \frac{1}{1 - R_j^2} \tag{8}$$

Nilai VIF lebih besar dari 10 menunjukkan adanya kolineritas antar variabel. Solusi untuk mengatasi adanya kasus multikolineritas adalah dengan mengeluarkan variabel prediktor yang tidak signifikasn dan meregresikan kembali variabel-variabel prediktor yang signifikan.

METODE PENELITIAN

E. Sumber Data dan Variabel Penelitian

Data dalam penelitian ini menggunakan data sekunder mengenai indikator pencemaran air secara biologi atau *Biocehmical Oxygen Demand* (BOD) yang diperoleh dari Badan Lingkungan Hidup Kota Surabaya pada periode tahun 2013. Pada penelitian ini memiliki unit observasi yang digunakan adalah 30 titik lokasi air badan sungai, dengan unit yang diteliti adalah sungai, saluran air, dan bozem yang ada di Surabaya. Data tersebut adalah jumlah kadar BOD dan faktor-faktor yang mempengaruhi sungai yang terdapat pada 30 titik lokasi Sungai Kota Surabaya.

Tabel 3.1 Variabel Penelitian

Variabel	Nama Variabel
Y	Biochemical Oxigen Demand (BOD) (mg/l)
x_1	Lebar sungai (Meter)
x_2	Kedalaman air sungai (Meter)
x_3	Kecepatan air sungai (m/detik)
x_4	Debit air sungai (m^3 detik)

F. Langkah Analisis Data

Dalam penelitian ini langkah-langkah yang dilakukan untuk memodelkan regresi gamma adalah sebagai berikut :

1. Memperoleh estimasi parameter untuk model regresi Gamma dengan metode MLE.
2. Melakukan pengujian hipotesis pada model Regresi Gamma dengan metode MLRT adalah sebagai berikut
3. Menentukan faktor yang berpengaruh terhadap kasus pencemaran kualitas air sungai di Surabaya menggunakan indikator BOD dengan mengaplikasikan model Reg resi Gamma.
   a. Melakukan analisis deskriptif varia bel respon (y) dan variabel prediktor (x).
   b. Melakukan pengujian kesesuaian dis tribusi Gamma terhadap data pada Y (BOD) dengan menggunakan uji An derson-Darling.
   c. Mengidentifikasi dan melakukan uji multikolineritas terhadap variabel independen menggunakan criteria VIF.

d. Mengaplikasikan model Gamma Regression pada pencemaran air sungai di Surabaya.
e. Melakukan pengujian signifikansi parameter pada model secara serentak dan parsial menggunakan MLRT.
f. Melakukan interpretasi model yang terbentuk.
g. Membuat kesimpulan dari hasil yang terbaik.

HASIL DAN PEMBAHASAN

Analisis dan pembahasan merupakan penyelesaian dari permasalahan yang ada. Hasil penelitian yang telah dilakukan dalam menjawab rumusan masalah yaitu membahas mengenai proses estimasi parameter dan bentuk statistik uji pada Regresi Gamma dengan menggunakan metode Maximum Likelihood Estimation (MLE) dan Weighted Least Square (WLS). Pada penelitian ini, selain kajian teoritis juga menggunakan data real yaitu data pencemaran air sungai yang dilakukan oleh Badan Lingkungan Hidup kota Surabaya tahun 2013 untuk menentukan faktor-faktor yang mempengaruhi pencemaran air sungai tersebut.

G. Estimator Parameter Regresi Gamma

Pada bagian ini membahas mengenai proses estimasi regresi Gamma dengan parameter skala θ merupakan fungsi dari variabel bebas atau kovariat, yakni $\theta = (\exp\,(\mathbf{x}_i{}^{\mathbf{T}}\boldsymbol{\beta})))/\alpha$,dengan $\boldsymbol{\beta} = (\beta_0\beta_1 \dots \beta_p)^T$, dan $\boldsymbol{x} = (1, X_{1\,,} \dots, X_i)^T$ serta diberikan fungsi kepadatan peluang bersama dari model regresi Gamma sebagai berikut.

$$f_{y_i}(y_i) = \frac{1}{\Gamma(\alpha)\left(\frac{\exp(\mathbf{x}_i{}^{\mathbf{T}}\boldsymbol{\beta})}{\alpha}\right)^{\alpha}} y_i^{\alpha-1} \exp\left(-y_i/\left(\frac{\exp(\mathbf{x}_i{}^{\mathbf{T}}\boldsymbol{\beta})}{\alpha}\right)\right)$$

$$\text{untuk } y_i > 0 \text{ dengan } \theta = \frac{\exp(\mathbf{x}^{\mathbf{T}}\boldsymbol{\beta})}{\alpha} \tag{9}$$

$$L(\boldsymbol{\beta}\,|\,\alpha) = \prod_{i=1}^{n} f_{y_i}(y_i)$$

$$= \prod_{i=1}^{n}\left[\frac{1}{\Gamma(\alpha)\left(\frac{\exp(\mathbf{x}_i{}^{\mathbf{T}}\boldsymbol{\beta})}{\alpha}\right)^{\alpha}} y_i^{\alpha-1}\exp\left(-y_i/\left(\frac{\exp(\mathbf{x}_i{}^{\mathbf{T}}\boldsymbol{\beta})}{\alpha}\right)\right)\right]$$

$$= \left(\frac{1}{\Gamma(\alpha)}\right)^{n}\left(\frac{\exp\sum_{i=1}^{n}\mathbf{x}_i{}^{\mathbf{T}}\boldsymbol{\beta}}{\alpha^n}\right)^{-\alpha}\left(\prod_{i=1}^{n} y_i^{\alpha-1}\right)$$

$$\exp\sum_{i=1}^{n}\left(-y_i/\left(\frac{\exp(\mathbf{x}_i{}^{\mathbf{T}}\boldsymbol{\beta})}{\alpha}\right)\right) \tag{10}$$

$$\ln L(\boldsymbol{\beta}\,|\,\alpha)$$

$$= \ln\left[\left(\frac{1}{\Gamma(\alpha)}\right)^{n}\left(\frac{\exp\sum_{i=1}^{n}\mathbf{x}_i{}^{\mathbf{T}}\boldsymbol{\beta}}{\alpha^n}\right)^{-\alpha}\left(\prod_{i=1}^{n} y_i^{\alpha-1}\right)\exp\sum_{i=1}^{n}\left(-y_i/\left(\frac{\exp(\mathbf{x}_i{}^{\mathbf{T}}\boldsymbol{\beta})}{\alpha}\right)\right)\right]$$

$$= -n\ln\Gamma(\alpha) - \alpha\sum_{i=1}^{n}\left(\mathbf{x_i}^{\mathbf{T}}\boldsymbol{\beta}\right) + \alpha(n\ln\alpha) + (n-1)\sum_{i=1}^{n}\ln y_i$$

$$-\sum_{i=1}^{n} y_i/\left(\exp\left(\mathbf{x_i}^{\mathbf{T}}\boldsymbol{\beta}\right)/\alpha\right) \tag{11}$$

Langkah selanjutnya adalah melakukan penaksiran parameter pada model regresi gamma dengan mencari turunan pertama secara parsial pada persamaan (11) terhadap masing-masing parameter yang diestimasi kemudian disamakan dengan nol.dan diperoleh turunan pertama fungsi lnlikelihood menghasilkan bentuk tidak closed form.

Turunan parsial pertama terhadap $\boldsymbol{\beta}$

$$\frac{\partial \ln L(\Omega)}{\partial \beta_0} = \alpha\sum_{i=1}^{n} y_i \exp\left(\mathbf{x}_i{}^{T}\boldsymbol{\beta}\right)^{-1} - \alpha n$$

$$\frac{\partial \ln L(\Omega)}{\partial \beta_1} = \alpha\sum_{i=1}^{n} x_{1i} y_i \left(\exp\left(\mathbf{x}_i{}^{T}\boldsymbol{\beta}\right)\right)^{-1} - \alpha\sum_{i=1}^{n} x_{1i}$$

$$\frac{\partial \ln L(\Omega)}{\partial \beta_2} = \alpha\sum_{i=1}^{n} x_{2i} y_i \left(\exp\left(\mathbf{x}_i{}^{T}\boldsymbol{\beta}\right)\right)^{-1} - \alpha\sum_{i=1}^{n} x_{2i}$$

$$\frac{\partial \ln L(\Omega)}{\partial \beta_3} = \alpha\sum_{i=1}^{n} x_{3i} y_i \left(\exp\left(\mathbf{x}_i{}^{T}\boldsymbol{\beta}\right)\right)^{-1} - \alpha\sum_{i=1}^{n} x_{3i}$$

$$\frac{\partial \ln L(\Omega)}{\partial \beta_4} = \alpha\sum_{i=1}^{n} x_{4i} y_i \left(\exp\left(\mathbf{x}_i{}^{T}\boldsymbol{\beta}\right)\right)^{-1} - \alpha\sum_{i=1}^{n} x_{4i}$$

Secara umum dapat dituliskan bentuknya seperti di bawah ini.

$$g(\boldsymbol{\beta}^{(m)}) = \left[\frac{\partial l}{\partial \boldsymbol{\beta}}\right]^T \qquad (12)$$

Oleh karena itu tidak dapat dianalisis secara analitik, untuk mendapatkan hasil yang eksplisit harus diselesaikan dengan metode numeric dengan iterasi Newton-Raphson untuk mendapatkan estimasi parameter. Persamaan iterasi Newton–Raphson dapat didefenisikan sebagai berikut.

$$\hat{\boldsymbol{\beta}}^{(m+1)} = \hat{\boldsymbol{\beta}}^{(m)} - H^{-1}(\hat{\boldsymbol{\beta}}^{(m)})g(\hat{\boldsymbol{\beta}}^{(m)}) \qquad (13)$$

Untuk menentukan persamaan iterasi Newton–Raphson diperlukan matriks Hessians. Matriks Hessian itu terdiri dari hasil turunan parsial kedua dari fungsi logaritma natural masing-masing parameter yang akan diestimasi dapat diperoleh :

$$\frac{\partial^2 \ln L(\Omega)}{\partial\beta_0\partial\beta_0} = -\alpha\sum_{i=1}^{n}\frac{y_i}{(\exp(\mathbf{x}_i^T\boldsymbol{\beta}))}$$

$$\frac{\partial^2 \ln L(\Omega)}{\partial{\beta_1}^2} = -\alpha\sum_{i=1}^{n}\frac{x_{1i}^2 y_i}{(\exp(\mathbf{x}_i^T\boldsymbol{\beta}))}$$

$$\frac{\partial^2 \ln L(\Omega)}{\partial{\beta_2}^2} = -\alpha\sum_{i=1}^{n}\frac{x_{2i}^2 y_i}{(\exp(\mathbf{x}_i^T\boldsymbol{\beta}))}$$

$$\frac{\partial^2 \ln L(\Omega)}{\partial{\beta_3}^2} = -\alpha\sum_{i=1}^{n}\frac{x_{3i}^2 y_i}{(\exp(\mathbf{x}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta}))}$$

$$\frac{\partial^2 \ln L(\Omega)}{\partial{\beta_4}^2} = -\alpha\sum_{i=1}^{n}\frac{x_{4i}^2 y_i}{(\exp(\mathbf{x}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta}))}$$

$$\frac{\partial^2 \ln L(\Omega)}{\partial\beta_1\partial\beta_0} = -\alpha\sum_{i=1}^{n}\frac{x_{1i} y_i}{(\exp(\mathbf{x}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta}))}$$

$$\frac{\partial^2 \ln L(\Omega)}{\partial\beta_2\partial\beta_0} = -\alpha\sum_{i=1}^{n}\frac{x_{2i} y_i}{(\exp(\mathbf{x}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta}))}$$

$$\frac{\partial^2 \ln L(\Omega)}{\partial\beta_3\partial\beta_0} = -\alpha\sum_{i=1}^{n}\frac{x_{3i} y_i}{(\exp(\mathbf{x}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta}))}$$

$$\frac{\partial^2 \ln L(\Omega)}{\partial\beta_4\partial\beta_0} = -\alpha\sum_{i=1}^{n}\frac{x_{4i} y_i}{(\exp(\mathbf{x}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta}))}$$

$$\frac{\partial^2 \ln L(\Omega)}{\partial\beta_0\partial\beta_1} = -\alpha\sum_{i=1}^{n}\frac{x_{1i} y_i}{(\exp(\mathbf{x}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta}))}$$

$$\frac{\partial^2 \ln L(\Omega)}{\partial\beta_2\partial\beta_1} = -\alpha\sum_{i=1}^{n}\frac{x_{1i}x_{2i} y_i}{(\exp(\mathbf{x}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta}))}$$

$$\frac{\partial^2 \ln L(\Omega)}{\partial\beta_3\partial\beta_1} = -\alpha\sum_{i=1}^{n}\frac{x_{1i}x_{3i} y_i}{(\exp(\mathbf{x}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta}))}$$

$$\frac{\partial^2 \ln L(\Omega)}{\partial\beta_4\partial\beta_1} = -\alpha\sum_{i=1}^{n}\frac{x_{1i}x_{4i} y_i}{(\exp(\mathbf{x}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta}))}$$

$$\frac{\partial^2 \ln L(\Omega)}{\partial\beta_0\partial\beta_2} = -\alpha\sum_{i=1}^{n}\frac{x_{2i} y_i}{(\exp(\mathbf{x}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta}))}$$

$$\frac{\partial^2 \ln L(\Omega)}{\partial\beta_1\partial\beta_2} = -\alpha\sum_{i=1}^{n}\frac{x_{1i}x_{2i} y_i}{(\exp(\mathbf{x}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta}))}$$

$$\frac{\partial^2 \ln L(\Omega)}{\partial\beta_3\partial\beta_2} = -\alpha\sum_{i=1}^{n}\frac{x_{2i}x_{3i} y_i}{(\exp(\mathbf{x}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta}))}$$

$$\frac{\partial^2 \ln L(\Omega)}{\partial\beta_4\partial\beta_2} = -\alpha\sum_{i=1}^{n}\frac{x_{2i}x_{4i} y_i}{(\exp(\mathbf{x}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta}))}$$

$$\frac{\partial^2 \ln L(\Omega)}{\partial\beta_0\partial\beta_3} = -\alpha\sum_{i=1}^{n}\frac{x_{3i} y_i}{(\exp(\mathbf{x}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta}))}$$

$$\frac{\partial^2 \ln L(\Omega)}{\partial\beta_1\partial\beta_3} = -\alpha\sum_{i=1}^{n}\frac{x_{1i}x_{3i} y_i}{(\exp(\mathbf{x}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta}))}$$

$$\frac{\partial^2 \ln L(\Omega)}{\partial\beta_2\partial\beta_3} = -\alpha\sum_{i=1}^{n}\frac{x_{2i}x_{3i} y_i}{(\exp(\mathbf{x}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta}))}$$

$$\frac{\partial^2 \ln L(\Omega)}{\partial\beta_4\partial\beta_3} = -\alpha\sum_{i=1}^{n}\frac{x_{3i}x_{4i} y_i}{(\exp(\mathbf{x}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta}))}$$

$$\frac{\partial^2 \ln L(\Omega)}{\partial\beta_0\partial\beta_4} = -\alpha\sum_{i=1}^{n}\frac{x_{4i} y_i}{(\exp(\mathbf{x}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta}))}$$

$$\frac{\partial^2 \ln L(\Omega)}{\partial\beta_1\partial\beta_4} = -\alpha\sum_{i=1}^{n}\frac{x_{1i}x_{4i} y_i}{(\exp(\mathbf{x}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta}))}$$

$$\frac{\partial^2 \ln L(\Omega)}{\partial\beta_2\partial\beta_4} = -\alpha\sum_{i=1}^{n}\frac{x_{2i}x_{4i} y_i}{(\exp(\mathbf{x}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta}))}$$

$$\frac{\partial^2 \ln L(\Omega)}{\partial\beta_3\partial\beta_4} = -\alpha\sum_{i=1}^{n}\frac{x_{3i}x_{4i} y_i}{(\exp(\mathbf{x}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta}))}$$

Secara umum dapat dituliskan bentuknya seperti di bawah ini.

$$H(\hat{\beta}^{(m)}) = \frac{\partial^2 \ln L(\Omega)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T} \qquad (14)$$

H. Uji Hipotesis Model Regresi Gamma

Pengujian hipotesisuntuk parameter Gamma menggunakan uji simultan dan uji parsial. Tujuan dari uji serentak adalah untuk mengkonfirmasi kovariat tertentu yang relevan. Hipotesis uji simultan berikut

$H_0 : \beta_1 = \cdots = \beta_k = 0$

$H_1 :$ paling sedikit ada satu $\beta_k \neq 0$, dengan j = $1,\cdots$,k.

Statistik pengujian untuk uji simultan ditentukan dengan metode MLRT, yang dilakukan dengan mencari set parameter di bawah H_0 dan di bawah populasi. Di bawah H_0, himpunan parameter diberikan oleh $\omega=\{\alpha,\theta,\beta_0\}$ dan fungsi likelihood nya diberikan oleh L(ω). Dibawah H_0, fungsi maksimum likelihood diberikan pada persamaan berikut.

$$L(\hat{\omega})=\prod_{i=1}^{n} f\left(\hat{\omega}\,|\,y_i\right)$$
$$=\left(\Gamma(\alpha_0)\theta^{\alpha_0}\right)^{-n}\prod_{i=1}^{n} y_i^{\alpha_0-1}\exp\sum_{i=1}^{n}\left(\frac{-y_i}{\theta}\right) \quad (15)$$

Fungsi logaritma natural dari fungsi likelihood bawah H_0 diberikan oleh persamaan 13

$$\ln L(\hat{\omega})=\ln\left[\left(\Gamma(\alpha_0)\theta^{\alpha_0}\right)^{-n}\prod_{i=1}^{n} y_i^{\alpha_0-1}\exp\sum_{i=1}^{n}\left(\frac{-y_i}{\theta}\right)\right]$$
$$=-n\Gamma(\alpha_0)\theta^{\alpha_0}+\left(\alpha_0-1\right)\sum_{i=1}^{n}\ln y_i+\sum_{i=1}^{n}\left(\frac{-y_i}{\theta}\right) \quad (16)$$

Sehingga diperoleh $L(\hat{\omega})=\max_{\omega} L\left(\omega\right)$ dengan masing-masing nilai parameter yang diperoleh dari iterasi Newton Raphson. Berikut dituliskan turunan parsial pertama untuk parameter dibawah H_0 yaitu parameter α_0,θ sama dengan 0 untuk memperoleh nilai setiap parameternya.

$$\frac{\partial \ln L(\omega)}{\partial \alpha_0}=-n\frac{1}{\Gamma(\alpha_0)}\Gamma'(\alpha_0)-n\ln\theta+\sum_{i=1}^{n}\ln y_i$$
$$\frac{\partial \ln L(\omega)}{\partial \theta}=-n\frac{(\alpha_0-1)}{\theta}+\sum_{i=1}^{n}\frac{y_i}{\theta^2}$$
$$\frac{\partial^2 \ln L(\omega)}{\partial {\alpha_0}^2}=-n\frac{\Gamma''(\alpha_0)\Gamma(\alpha_0)-\Gamma'(\alpha_0)\Gamma'(\alpha_0)}{\Gamma^2(\alpha_0)}$$
$$\frac{\partial^2 \ln L(\omega)}{\partial \theta^2}=-n\frac{1}{\theta^2}(\alpha_0-1)-2\sum_{i=1}^{n}\frac{y_i}{\theta^3}$$
$$\frac{\partial^2 \ln L(\omega)}{\partial\theta\partial\alpha_0}=\frac{-n}{\theta}$$
$$\frac{\partial^2 \ln L(\omega)}{\partial\alpha_0\partial\theta}=\frac{-n}{\theta}$$

Dari hasil penurunan parsial untuk turunan kedua dari fungsi logaritma natural likelihood model regresi Gamma diperoleh matriks $H(\omega^{(m)})$ sebagai berikut.

$$H(\omega^{(m)})=\begin{bmatrix}\frac{\partial^2 \ln L(\omega)}{\partial\alpha_0^2} & \frac{\partial^2 \ln L(\omega)}{\partial\theta\partial\alpha_0}\\ \frac{\partial^2 \ln L(\omega)}{\partial\alpha_0\partial\theta} & \frac{\partial^2 \ln L(\omega)}{\partial\theta^2}\end{bmatrix}$$

Himpunan dari parameter di bawah populasi diberikan pada Ω = {α, θ, β_0, β_1, β_2, β_3, β_4}. Fungsi likelihood di bawah populasi (H_1) adalah L(Ω), dimana L(**β**|α) dari persamaan (7). population Fungsi maksimum likelihood di bawah populasi $L(\hat{\Omega})=\max_{\Omega} L(\boldsymbol{\beta}\,|\,\alpha)$ dan fungsi logaritma natural maksimum likelihood diberikan pada persamaan 16 :

$$L(\hat{\Omega})=\max_{\Omega} L(\boldsymbol{\beta}\,|\,\alpha) \quad (17)$$

Untuk menentukan uji statistik pada uji serentak adalah dengan metode Likelihood Ratio Test, dengan statistik uji :

$$G^2=-2\ln\Lambda=-2\ln\frac{L(\hat{\omega})}{L(\hat{\Omega})}=-2\ln\left(L(\hat{\Omega})-L(\hat{\omega})\right)$$
$$=2\ln L(\hat{\Omega})-2\ln L(\hat{\omega}) \quad (18)$$
$$G^2=\left(\begin{array}{l}2\ln n\Gamma(\alpha_0)\theta^{\alpha_0}-2\left(\alpha_0-1\right)\sum_{i=1}^{n}\ln y_i-2\sum_{i=1}^{n}\left(\frac{-y_i}{\theta}\right)+\\ -2n\ln\Gamma\left(\alpha\right)-2\alpha\sum_{i=1}^{n}\left(x_i^T\beta\right)+2\alpha(n\ln\alpha)\\ +2(n-1)\sum_{i=1}^{n}\ln y_i-2\sum_{i=1}^{n}y_i/\left(\exp\left(x_i^T\beta\right)/\alpha\right)\end{array}\right)$$

Dimana dapat diturunkan likelihood rasio statistik:

$$LR=\frac{L(\hat{\omega})}{L(\hat{\Omega})}$$

Di bawah H0, uji statistik G asymptotic dengan distribusi χ^2 dengan derajat bebas 2p dan tingkat signifikan α dimana 0 < α < 1, jika nilai dari G melebihi sesuai dengan (1-α) –kuantil dari distribusi χ^2 maka tolak H_0.

Uji parsial untuk parameter regresi Gamma bertujuan untuk mengetahui parameter yang memberikan pengaruh signifikan terhadap model. Hipotesis uji parsial :

$$H_0:\beta_j=0$$
$$H_1:\ \beta_j\neq 0,\ \text{with j = 1, 2,}\cdots\text{, k}$$

Uji statistik dari pengujian parameter regresi dengan uji parsial $H_0 : \hat{\beta}_j = 0$ diberikan statistik Wald :

$$Z = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad (19)$$

dengan Z ~ N(0,1), $SE(\hat{\beta}_j) = \sqrt{\text{var}(\beta_j)}$ dan $\text{var}(\hat{\beta}_j)$ adalah elemen diagonal utama ke-(j+1) matriks Hessian $-\left[H(\hat{\beta})\right]^{-1}$, dengan j = 0, 1, 2, …, k. Kriteria penolakan H_0 apabila $\left|Z_{hitung}\right| > Z_{tabel}$

I. Faktor-faktor yang Mempengaruhi Tingginya Nilai BOD Sungai di Surabaya

Analisis model regresi Gamma pada penelitian ini diaplikasikan pada kasus pencemaran air sungai di Surabaya pada tahun 2013. Data yang diperoleh dari Badan Lingkungan Hidup (BLH) digunakan untuk mengetahui faktor-faktor yang mempengaruhi pencemaran sungai di Surabaya berdasarkan kadar BOD. Sebelum melakukan analisis faktor yang mempengaruhi pencemaran air sungai di Surabaya terlebih dahulu mendeskripsikan karakteristik dari data zat pencemar dari data BLH.

Tabel 2 Estimasi Parameter Model Regresi Gamma dengan Estimasi MLE Program Matlab

Parameter	Program Matlab			
	Estimasi	Zhitung	p-value	Kesimpulan
β_0	2,5210	20,8110	0	Tolak H_0
β_1	-0,0016	-0,3055	0,3800**	Gagal Tolak H_0
β_2	-0,0907	-1,4391	0,0751*	Gagal Tolak H_0
β_3	-0,1941	-1,6738	0,0471	Tolak H_0
β_4	0,0002	0,6641	0,2533*	Gagal Tolak H_0

Keterangan : *)signifikan=10%; **)signifikan = 30% ; ***)signifikan = 40%

Tabel 2 menunjukkan bahwa keempat variabel yang berpengaruh terhadap kadar Biochemical Oxygen Demand (BOD) di Kali Surabaya untuk Lebar Sungai (x_1) berpengaruh ketika tingkat signifikansinya sebesar 40 %, variabel kedalaman Sungai (x_2) signifikan pada 10% dan variabel Kecepatan Air (x_3) yang hanya signifikan pada α=0,05. Hal ini menandakan bahwa Kedalaman Sungai dan kecepatan air sungai begitu mempengaruhi kadar Biochemical Oxygen Demand berdasarkan signifikansi yang mempengaruhinya. Sedangkan untuk Debit Sungai (x_4) signifikan ketika tingkat signifikansinya 30 %. Hal ini menunjukkan bahwa debit sungai itu tidak cukup mempengaruhi kadar Biochemical Oxygen Demand.

Dengan demikian, variabel yang paling signifikan mempengaruhi faktor-faktor pencemaran sungai di Surabaya dengan indikator BOD adalah Kecepatan Air Sungai (x_3). Sehingga salah satu kebijakan yang diambil untuk mengantisipasi pencemaran air di sungai kota Surabaya dengan melakukan pengorekan ke badan-badan sungai sehingga alirannya lancar dan dapat mencegahnya pencemaran sungai dan sekaligus mengurangi pendangkalan bantaran kalinya.

Tabel 3 Estimasi Parameter Model Regresi Gamma dengan Estimasi WLS dengan Program Matlab

Parameter	Program MATLAB			
	Estimasi	zhitung	p-value	Kesimpulan
β_0	11,95	13,535	0,0000	Tolak H_0
β_1	-0,0108	-1,590	0,1155 *	Gagal Tolak H_0
β_2	-0,4691	-1,503	0,1365 *	Gagal Tolak H_0
β_3	-2,1182	-3,582	0,0006	Tolak H_0
β_4	0,0019	1,332	0,1864 **	Gagal Tolak H_0

Keterangan : *)signifikan = 15% ; **)signifikan = 20%

Tabel 3 menunjukkan bahwa keempat variabel yang berpengaruh terhadap kadar Biochemical Oxygen Demand (BOD) di Kali Surabaya pada tingkat signifikansi 5% adlah

Kecepatan Air (x_3). Untuk Lebar Sungai (x_1) dan Kedalaman Sungai (x_2) berpengaruh ketika tingkat signifikansinya sebesar 15%, sedangkan variabel Debit Air Sungai (x_4) berpengaruh ketika tingkat signifikansinya 20%.

Model dengan menggunakan metode estimasi MLE dihasilkan dari pengolahan data menggunakan program matlab adalah sebagai berikut.

$$\hat{\theta} = \frac{\exp(2,5210 - 0,0016x_1 - 0,0907x_2 - 0,1942x_3 + 0,002x_4)}{3,7485}$$

Pada model diatas menjelaskan setiap perubahan 1 m/detik kecepatan sungai akan melipatkangandakan nilai parameter θ sebesar exp (-0,194/3,7485)= 0,95 g/l kadar BOD dengan asumsi kadar lain dianggap konstan.

Model regresi Gamma dengan menggunakan metode estimasi *Weighted Least Square* (WLS) dihasilkan dari pengolahan data menggunakan program matlab adalah sebagai berikut.

$$\hat{\theta} = \frac{\exp(11,95 - 0,0108x_1 - 0,4691x_2 - 2,1182x_3 + 0,0019x_4)}{3,7485}$$

Pada model diatas menjelaskan setiap perubahan 1 m/detik kecepatan sungai akan melipatkangandakan nilai parameter θ sebesar exp(-2,118/3,7485)= 0,57 g/l kadar BOD dengan asumsi kadar lain dianggap konstan.

Sebagai perbandingan antara model regresi Gamma menggunakan metode estimasi maksimum likelihood dan weighted least square.

Tabel 4. Nilai AIC Model

Model	Nilai AIC
Maximum Likelihood Estimation	4,22
Weighted Least Square	0,68

Berdasarkan tabel 4 menunjukkan bahwa nilai AIC untuk metode estimasi Weighted Least Square (WLS) menghasilkan nilai AIC yang lebih kecil dibandingkan dengan metode estimasi Maximum Likelihood. Oleh karena itu dapat disimpulkan bahwa metode estimasi *Weighted Least Square* yang dipilih untuk pemodelan pencemaran sungai di Surabaya dengan indikator *Biochemical Oxygen Demand* (BOD).

V. KESIMPULAN DAN SARAN

A. Kesimpulan

Dalam penelitian ini difokuskan pada model regresi Gamma, yang digunakan untuk data kontinu non-negatif. Metode analisis untuk estimasi parameter adalah MLE, yang didasarkan pada rata-rata dari distribusi Gamma. Estimasi maksimum likelihooddapat diperoleh dengan menyelesaikan sistem linear yangtidaksaling bergantung untuk turunan parsial pertama dari log likelihood sama dengan nol, menggunakan iterasi pengulangan Newton-Raphson untuk mendapatkan estimatornya. Untuk estimasi parameter dibawah populasi menghasilkan 7 parameter yaitu $\Omega = \{\alpha, \theta, \beta_0, \beta_1, \beta_2, \beta_3, \beta_4\}$ sedangkan untuk estimasi parameter dibawah H0 menghasilkan 3 parameter yaitu $\omega = (\beta_0, \alpha, \theta)$.

Pengujian hipotesis untuk parameter regresi dalam model regresi Gamma melibatkan uji kebaikan fit, secara uji serentak dan uji parsial. Uji statistik untuk kebaikan fit dan uji serentakmenggunakan statistik Wilk likelihood rasio, yang ditentukan dengan metode LRT, dan pengujian statistik untuk uji parsial adalah dengan menggunakan statistik Wald likelihood rasio, turunan dari sifat asymptotic estimasi maksimum likelihood.

Selanjutnya, untuk studi lebih lanjutnya adalah mengenai konvergensi dari algoritma iterasi Newton-Raphson untuk memperoleh nilai awal dan estimasi yang lebih tepat dalam model regresi Gamma sehingga model yang diperoleh efisiensi dari metode ini. Pada pengujian hipotesis secara serentak pada metode MLE dan WLS menghasilkan terdapat minimal satu variabel prediktor yang mempengaruhi terhadap variabel respon, sehingga dilanjutkan ke pengujian parsial. Pada pengujian hipotesis parsial metode MLE dan WLS menunjukkan variabel prediktor yang berpengaruh terhadap variabel respon adalah kecepatan air sungai.

B. Saran

Berdasarkan penelitian yang telah dilakukan, permasalahan yang dikaji dapat dikembangkan untuk penelitian-penelitian berikutnya adalah menambahkan indikator lain untuk variabel respon seperti Chemical Oxygen Demand (COD) dan Dissolved Oxygen (DO). Selanjutnya dapat dapat ditambahi dengan efek spasial sehingga dapat menggunakan Geographically Weighted

Gamma Regression.

DAFTAR PUSTAKA

Sebah, Pascal., dan Gourdon, Xavier. (2002). *Introduction to the Gamma Function.*

T. Izawa, Two or multi-dimensional gamma-type distribution and its application to rainfall data. *Papers in Meteorology and Geophysics*, **15** (1965), 167–200.

R.T. Clarke, Bivariate gamma distribution for extending annual streamflow record from precipitation: Some large-sample results. *Water Resources Research*,16 (1980), 863–870.

Mathai A.M. and Moschopoulos P.G. On A Multivariate Gamma (1991). *Journal of Multivariate Analysis*, **Vol. 39, No 1**, pp. 135-153.

Williams, E. J., (1959). *Regression Analysis*, John Wiley & Sons, New York.

Al-Abood, Akram. (1986). The Power of Approximate Test for the Regression Coefficients in a Gamma Regression Model. *IEEE Transaction on Reliability*, 35(2).

Schhutz, Aurelien. & Lionel, Bombrun. (2013). Centroid-Based Texture Classification Using The Generalized Gamma Distribution. *HAL*, 00878744.

McCullagh, P. & Nelder, J. A. (1989). *Generalized Linear Model 2nd Edition*. London: Chapman & Hall.

Casella, G.. & Berger, R. L. (2002). *Statistical Inference* (2 ed.): Duxbury.

Montgomery, D.C., Runger, G.C. (2003). *Applied Statistics and Probability for Engineer*, 3rd ed.. John Wiley & Son, Inc.

Ghozali, Imam.(2005). *Aplikasi Analisis Multivariate dengan program SPSS*. Badan Penerbit Universitas Diponegoro, Semarang

Mattjik, A. A & Sumertajaya, I. M. 2000. Perancangan Percobaan dengan Aplikasi SAS dan Minitab Jilid I. Bogor: IPB Press.