PENDETEKSIAN *OUTLIER* PADA REGRESI NONLINIER DENGAN METODE STATISTIK *LIKELIHOOD DISPLACEMENT* (LD)

Siti Tabi'atul Hasanah

Mahasiswa Jurusan Matematika UIN Maulana Malik Ibrahim Malang
e-mail: chib_y@yahoo.com

ABSTRAK

Outlier merupakan pengamatan yang jauh berbeda (ekstrim) dari data pengamatan lainnya, atau dapat diartikan data yang tidak mengikuti pola umum model. Adakalanya *outlier* memberikan informasi yang tidak dapat diberikan oleh data yang lainnya. Karena itulah outlier tidak boleh begitu saja dihilangkan. *Outlier* dapat juga merupakan pengamatan berpengaruh. Banyak sekali metode yang dapat digunakan untuk mendeteki adanya *outlier.* Pada penelitian-penelitian sebelumnya pendeteksian *outlier* dilakukan pada regresi linier. Selanjutnya akan dikembangkan pendeteksian *outlier* pada regresi nonlinier. Regresi nonlinier disini dikhususkan pada regresi nonlinier multiplikatif. Untuk mendeteksi yaitu menggunakan metode statistik *likelihood displacement*. Metode statistik *likelihood displacement* disingkat *(LD)* adalah suatu metode untuk mendeteksi adanya outlier dengan cara menghilangkan data yang diduga *outlier*. Untuk mengestimasi parameternya maka digunakan metode *maximum likelihood*, sehingga didapatkan hasil etimasi yang maksimal. Dengan metode LD diperoleh LD_{A_k}, yaitu *likelihood displacement* yang diduga mengandung *outlier*. Selanjutnya Keakuratan metode LD dalam mendeteksi adanya *outlier* ditunjukkan dengan cara membandingkan MSE dari LD dengan MSE dari regresi pada umumnya. Statistik uji yang digunakan adalah Λ. Hipotesis awal ditolak ketika $\chi^2_{hitung} > \chi^2_{tabel}$, sehingga terbukti LD_{A_k} adalah suatu *outlier.*

Kata kunci: *likelihood displacement, maximum likelihood estimation, outlier, regresi nonlinier multiplikatif.*

ABSTRACT

Outlier is an observation that much different (extreme) from the other observational data, or data can be interpreted that do not follow the general pattern of the model. Sometimes outliers provide information that can not be provided by other data. That's why outliers should not just be eliminated. Outliers can also be an influential observation. There are many methods that can be used to detect of outliers. In previous studies done on outlier detection of linear regression. Next will be developed detection of outliers in nonlinear regression. Nonlinear regression here is devoted to multiplicative nonlinear regression. To detect is use of statistical method likelihood displacement. Statistical methods abbreviated likelihood displacement (LD) is a method to detect outliers by removing the suspected outlier data. To estimate the parameters are used to the maximum likelihood method, so we get the estimate of the maximum. By using LD method is obtained LD_{A_k} i.e likelihood displacement is thought to contain outliers. Further accuracy of LD method in detecting the outliers are shown by comparing the MSE of LD with the MSE from the regression in general. Statistic test used is Λ. Initial hypothesis was rejected when $\chi^2_{count} > \chi^2_{table}$, proved so LD_{A_k} is an outlier.

Keywords: likelihood displacement, maximum likelihood estimation, multiplicative nonlinear regression, Outlier

PENDAHULUAN

Outlier adalah pengamatan yang jauh berbeda (ekstrim) dari data pengamatan lainnya. Salah satu penyebab terjadinya *outlier* adalah kesalahan pada pengambilan data sehingga menyebabkan data tersebut menjadi ekstrim. Adakalanya *outlier* ini tidak boleh begitu saja dihilangkan, namun dalam hal ini harus hati-hati karena terkadang *outlier* itu memberikan informasi yang tidak dapat diberikan oleh titik pengamatan lain, misalnya karena adanya kombinasi keadaan yang tidak biasa dan perlu diadakan penyelidikan lebih jauh. Suatu *outlier* dapat dibuang setelah ditelusuri ternyata pengamatan tersebut merupakan akibat dari kesalahan pengukuran atau kesalahan dalam menyiapkan pengukuran. *Outlier* dapat juga merupakan pengamatan berpengaruh. *Outlier* yang bukan pengamatan berpengaruh, tidak memiliki pengaruh yang kuat pada model kecuali *outlier* tersebut sangat besar. Tetapi jika *outlier* merupakan data berpengaruh, maka akan memberikan dampak pada model (Drapper dan Smith, 1992:146).

Misalkan saja pada suatu penelitian tentang sapi penghasil susu. Dari suatu data ternyata diperoleh ada beberapa sapi yang menghasilkan hasil susu yang lebih banyak dari biasanya atau dari sapi normalnya. Sapi penghasil susu yang tidak sesuai dengan normalnya merupakan suatu *outlier,* namun jika mengahapus begitu saja data ini berarti telah menghilangkan bibit sapi unggul yang mampu menghasilkan banyak susu sapi. Oleh sebab itulah penting untuk mengidentifikasi adanya *outlier* agar tidak kehilangan suatu data yang memiliki kualitas yang bagus. Jika dengan adanya *outlier* itu kurang baikmaka perlu diidentifikasi dan kemudian dihilangkan data yang mengandung *outlier.*

Banyak sekali metode yang dapat digunakan untuk mendeteksi *outlier*, salah satunya yaitu pendektesian *outlier* pada model linier univariat telah dikemukaan oleh Cook dengan memperkenalkan Jarak Cook (*Cook's Distance*) sebagai ukuran untuk mendeteksi pengamatan berpengaruh dalam model linier univariat. Ukuran Jarak Cook ini dirumuskan sebagai kombinasi dari *studential residual,* variansi *residual,* dan variansi nilai prediksi. Selain metode yang dikemukakan oleh Cook, masih banyak lagi metode yang digunakan untuk pendeteksian *outlier* pada model linier (Makkulau, 2010:95)

Xu, Abraham dan Steiner (2005) mengembangkan Jarak Cook univariat untuk mendeteksi *outlier* pada model linier multivariat (model regresi linier multivariat) dengan menggunakan metode statistik *likelihood displacement* yang disingkat LD. Metode LD adalah suatu metode untuk mendeteksi adanya *outlier* dengan cara menghilangkan pengamatan yang diduga *outlier* (Makkulau, 2010:95).

Tujuan dari penelitian ini adalahuntuk mengetahui cara mendeteksi *outlier* pada regresi nonlinier dengan metode statistik *Likelihood Displacement* (LD).

Mafaat dari penelitian ini adalah untuk mengembangkan metode yang dapat digunakan untuk mendeteksi adanya *outlier.*

KAJIAN TEORI

1. *Outlier*

Secara umum *outlier* dapat diartikan data yang tidak mengikuti pola umum pada model atau data yang keluar dari model dan tidak berada dalam daerah selang kepercayaan (Sembiring, 1995:62).

Menurut Draper dan Smith (1992:146) sisaan yang merupakan *outlier* adalah yang nilai mutlaknya jauh lebih besar dari pada sisaan lainnya dan terletak tiga atau empat kali simpangan baku atau lebih jauh lagi dari rata-rata sisaannya. *Outlier* merupakan suatu keganjilan dan menandakan suatu titik data yang sama sekali tidak tipikal dibandingkan data lainnya.

2. Estimasi Parameter

Menurut Yitnosumarto (1990:211) penduga (*estimator*) adalah anggota peubah acak statistik yang mungkin untuk sebuah parameter (anggota peubah yang diturunkkan). Parameter adalah nilai yang mengikuti acuan keterangan atau informasi yang dapat menjelaskan batas-batas atau bagian-bagian tertentu dari suatu sistem persamaan.

Murray dan Larry (1999:166) menyatakan terdapat dua jenis estimasi parameter, yaitu: estimasi titik dan estimasi interval.

Estimasi titik adalah Estimasi dari sebuah parameter populasi yang dinyatakan oleh bilangan tunggal disebut sebagai estimasi titik dari parameter tersebut. Sebuah nilai yang diperoleh dari sampel dan digunakan sebagai estimasi dari parameter yang nilainya tidak diketahui. Misalkan $X1, X2, \dots, Xn$ merupakan sampel acak berukuran n dari X, maka statistik yang berkaitan dengan θ dinamakan estimasi dari θ. Setelah sampel diambil, nilai-nilai yang dihitung dari sampel itu digunakan sebagai taksiran titik bagi θ.

Estimasi dari parameter populasi yang dinyatakan dengan dua bilangan. Di antara posisi parameternya diperkirakan berbeda, sehinggga disebut estimasi interval. Estimasi interval mengindikasikan adanya tingkat kepresisian atau akurasi dari sebuah estimasi sehingga estimasi interval akan dianggap semakin baik jika mendekati estimasi titik

Adapun sifat-sifat estimasi titik adalah sebagai berikut:

1. Tak Bias

   Yusuf Wibisono (2005:362) dalam bukunya menyatakan bahwa *estimator* tak bias bagi parameter θ, jika
   $$E(\hat{\theta}) = \theta$$
2. Konsisten

   Damodar N. Gujarati (2007:98) menerangkan *estimator* parameter $\hat{\theta}$ dikatakan konsisten bila nilai-nilainya mendekati nilai parameter yang sebenarnya meskipun ukuran sampelnya semakin besar.
3. Efisien

   Jika distribusi sampling dari dua statistik memiliki *mean* atau ekspektasi yang sama, maka statistik dengan variansi yang lebih kecil disebut sebagai *estimator* efisien dari *mean*, sementara statistik yang lain disebut *estimator* tak efisien. Adapun nilai-nilai yang berkorespondensi dengan statistik-statistik ini masing-masing disebut sebagai estimasi efisien dan estimasi tak efisien.

4. Distribusi

Suatu peubah acak X berdistribusi normal $N(\mu, \sigma^2)$ bila untuk suatu $\sigma^2 > 0$ dan $-\infty < \mu < \infty$ (Turmudi dan Harini, 2008:204).

mempunyai fungsi densitas pada $X = x$ dengan persamaan:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad (2.1)$$

Distribusi lain yang digunakan yaitu distribusi chi-square. Distribusi *chi-square* merupakan distribusi dengan variabel acak kontinue. Simbol untuk *chi-square* adalah χ^2. Distribusi *chi-square* sebenarnya merupakan jumlah kuadrat dari variabel-variabel acak yang bebas dan menyebar normal dengan *mean* 0 dan ragam $1\big(Z \sim NID(0,1)\big)$. Distribusi ini dapat dinyatakan dengan

$$X^2 = Z_1^2 + Z_2^2 + \cdots + Z_k^2$$

$$\chi^2 = \sum_i Z_i^2 \sum_i \left(\frac{Y_i - \mu_i}{\sigma_i}\right)^2$$

merupakan variabel acak yang tersebar menurut distribusi *chi-square* dengan derajat bebas sebesar k dan dapat dituliskan

$$X^2 \approx \chi_k^2$$

dimana χ_k^2 yaitu distribusi *chi-square* dengan derajat bebas k.

Suatu variabel acak X berdistribusi *chi-square* dengan derajat bebas k, dinyatakan dengan $\chi_k^2(0)$ bila untuk suatu bilangan bulat $k > 0$. (Turmudi dan Harini, 2008: 210)

Distribusi ini mempunyai fungsi kepekatan peluang sebagai berikut:

$$f_x(x) = \begin{cases} \dfrac{1}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)} x^{\left(\frac{k}{2}\right)-1} e^{-\frac{x}{2}}, & x \geq 0 \\ 0, & \text{selainnya} \end{cases}$$

Nilai tengah (*mean*) dan ragam untuk distribusi χ^2 adalah $\mu = k$ dan $s^2 = 2k$. Distribusi *chi-square* bergantung pada banyaknya simpangan baku yang bebas antara satu dengan yang lain atau dengan kata lain bergantung pada derajat bebasnya.

Jika X dan Y variabel acak, maka peluang terjadinya X dan Y secara serentak dinyatakan sebagai $f(x, y)$ disebut Distribusi Peluang Gabungan untuk setiap pasangan (x, y) (Herrhyanto, 2009:5).

5. Regresi Nonlinier

Analisis regresi merupakan analisis yang menyangkut studi tentang hubungan antara satu variabel yang disebut variabel terikat atau variabel yang dijelaskan dan satu atau lebih variabel yang lain yang disebut variabel bebas atau variabel penjelas (Gujarati, 2007:115).

Regresi yang variabel-variabelnya berbentuk tidak biasa. Bentuk grafik regresi nonlinier adalah berupa lekungan (Hasan, 2002:297).

Model regresi nonlinier dapat digolongkan menjadi dua yaitu model linier *intrinsik* dan model nonlinier *intrinsik*. Jika suatu model dikatakan model linier *intrinsik*, maka model model ini dapat dinyatakan dalam bentuk linier baku dengan mentransformasikan secara tepat terhadap peubahnya. Jika suatu model nonlinier tidak dapat dinyatakan dalam bentuk baku, berarti model ini secara intsinsik adalah nonlinier. Berikut ini adalah beberapa model yang dapat dinyatakan dalam linier baku (Draper dan Smith, 1992:213).

6. Regresi Multiplikatif

Regresi Multiplikatif adalah salah satu bentuk dari regresi linier *intrinsik*. Bentuk umum dari regresi multiplikatif adalah sebagai berikut:

$$Y = \propto X_1^{\beta} X_2^{\delta} X_3^{\delta} \varepsilon \qquad (2.2)$$

dimana σ, β, dan δ adalah parameter yang tidak diketahui, dan ε adalah galat acak yang bersifat multiplikatif. Dengan mengalgoritmakan basis e pada pada persamaan di atas, maka model persamaan di atas menjadi $ln\ Y = ln\,\alpha + \beta\, ln\, X_1 + \gamma\, ln\, X_2 + \delta\, ln\, X_3 + ln\,\varepsilon$.

Model persamaan tersebut menjadi bentuk linier sehingga dapat ditangani dengan prosedur regresi nonlinier. Model tersebut merupakan model linier dalam bentuk $ln\,\varepsilon$. ε tidak berdistribusi normal, sebab yang berdistribusi normal adalah $ln\,\varepsilon$ (Draper dan Smith, 1992:213).

7. Regresi dalam Pendekatan Matriks

Model regresi yang paling sederhana adalah model regresi linier. Model regresi linier sederhana terdiri dari satu variabel. Model tersebut dapat digeneralisasikan menjadi lebih dari satu atau dalam *k* variabel. Persamaan model regresi linier dengan k peubah adalah sebagai berikut:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon \qquad (2.3)$$

pengamatan mengenai $y, x_1, x_2, \ldots, x_k$ dinyatakan masing-masing dengan $y_i, x_{i1}, x_{i2}, \ldots, x_{ik}$ dan galatnya ε_i, maka persamaan (2.3) dapat dituliskan sebagai:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_k + \varepsilon_i$$

untuk, $i = 1,2,\ldots,n$. Dinotasikan dalam bentuk matriks, sehingga menjadi:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Misalkan

$$\underline{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix}$$

$$\underline{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} \quad \underline{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Persamaan (2.11) dapat dinyatakan sebagai:

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon} \qquad (2.4)$$

dengan:

$\underline{Y}$: vektor respon n x 1

X :matriks peubah bebas berukuran n x $(k+1)$

$\underline{\beta}$: vektor parameter berukuran $(k+1)$ x 1

$\underline{\varepsilon}$: vektor galat ukuran n x 1

(Sembiring,1995:134-135)

8. *Maximum likelihood*

Statistik inferensia dapat dibagi dalam dua bagian besar, estimasi dan pengujian hipotesis. Kedua inferensi tersebut masing-masing bertujuan untuk membuat pendugaan dan pengujian suatu parameter populasi dan informasi sampel yang diambil dari populasi tersebut. Gujarati N. Damodar (2010:131) menjelaskan bahwa metode dari estimasi titik (*point estimation*) dengan sifat-sifat teoritis yang lebih kuat dari pada metode OLS adalah metode *maximum likelihood* (ML).

Fungsi *likelihood* dari n peubah acak $x_1, x_2, \dots, x_n$ didefinisikan sebagai fungsi kepadatan bersama dari n peubah acak. Fungsi kepadatan bersama $f_{x_1},\dots_{x_n}(x_1, \dots, x_n; \theta)$, yang mempertimbangkan fungsi dari θ. Jika $x_1, \dots, x_n$ adalah sampel acak dari fungsi kepadatan $f(x,\theta)$, maka fungsi *likelihood*nya adalah $f(x_1;\theta)f(x_2;\theta)\dots f(x_n;\theta)$ (Mood, Graybill and Boes, 1986:278).

Maximum likelihood dapat diperoleh dengan menentukan turunan dari L terhadap parameternya dan menyatakannya sama dengan nol. Dalam hal ini, akan lebih mudah untuk terlebih dahulu menghitung logaritma kemudian menentukan turunannya. Dengan cara ini diperoleh:

$$\frac{1}{f(X_1,\theta)}\frac{\partial f(x_1,\theta)}{\partial \theta} + \cdots + \frac{1}{f(X_n,\theta)}\frac{\partial f(x_n,\theta)}{\partial \theta} = 0$$

Penyelesaian dari persamaan ini, untuk θ dalam bentuk x_k, dikenal sebagai estimator *maximum likelihood* dari θ.

9. Metode Statistik *Likelihood Displacement* (LD)

Metode LD adalah suatu metode yang dikembangkan dengan cara menghilangkan pengamatan yang diduga *outlier*. Misalkan k adalah pengamatan dikumpulkan pada pengamatan tertentu, dengan k diduga sebagai *outlier*. Indeks A_k adalah kumpulan dari k yang diduga *outlier.*

LD dari pengamatan yang mengandung *outlier* untuk $\hat{\beta}^*$ dengan variansi $\hat{\sigma}^2$ adalah:

$$LD_{A_k}\left(\hat{\beta}^* \middle| \hat{\sigma}^2\right) = 2\left\{ln\, L\left(\hat{\beta}^*, \hat{\sigma}^2\right) - ln\, L\left(\hat{\beta}^*_{[A_k]}, \hat{\sigma}^2(\hat{\beta}^*_{[A_k]})\right)\right\} \qquad (2.5)$$

dimana $\hat{\sigma}^2(\hat{\beta}^*_{[A_k]})$ adalah MLE dari $\hat{\sigma}^2$ ketika $\hat{\beta}^*$ diestimasi oleh $\hat{\beta}^*_{[A_k]}$ (Makkulau, dkk, 2010:97).

PEMBAHASAN

1. Regresi Nonlinier Multiplikatif

Bentuk umum dari regresi nonlinier multiplikatif adalah dinyatakan sebagai berikut:

$$y_i = \beta_0 x_{i1}^{\beta_1} x_{i2}^{\beta_2} x_{i3}^{\beta_3} \dots x_{nk}^{\beta_k} \dots \varepsilon_i \qquad (3.1)$$

Persamaan (3.1) dapat dilinierkan dengan melogaritmanaturalkan persamaannya, sehingga modelnya menjadi:

$$ln\, y_i = ln\, \beta_0 + \beta_1\, ln\, x_{1i} + \beta_2\, ln\, x_{2i} + \cdots + \beta_k\, ln\, x_{nk} + \cdots + ln\, \varepsilon_i \qquad (3.2)$$

dengan $i = 1,2, \dots, n$ dan $k = 1,2, \dots, n$

Dalam penelitian ini diasumsikan bahwa variabel terikat $(ln\, y)$ berdistribusi normal dengan *mean* μ dan variansi σ^2. Sehingga dalam persamaan (3.1) ε berdistribusi log normal, karena yang berdistribusi normal adalah ln ε.

Dengan menggunakan pendekatan matriks, diperoleh:

$$Y^*_{n\,\text{x}\,1} = X^*_{n\,\text{x}(k+1)}\beta^*_{(k+1)\text{x}\,1} + \varepsilon^*_{n\,\text{x}\,1} \qquad (3.3)$$

2. Estimasi parameter regresi noninier multiplikatif

Dari persamaan (3.3) diketahui bahwa $Y^* = (ln\, y_1, ln\, y_2, \dots, ln\, y_n)^T$ adalah variabel random, karena diasumsikan berdistribusi normal, maka $Y^* \sim N(X^*\beta^*, I\sigma^2)$ dengan $X^* = (ln\, x_{0i}, ln\, x_{1i}, \dots, ln\, x_{1k})$ dan $\beta^* = (ln\, \beta_0, \beta_1, \dots, \beta_n)^T$ dimana $i = 1,2, \dots, n$ dan I adalah matriks identitas. Sehingga fungsi distribusi peluang gabungannya adalah

$$f(Y^*|\beta^*, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}(Y^*-X^*\beta^*)^T(Y^*-X^*\beta^*)}$$

(3.4)

sehinggga fungsi likelihoodnya adalah:

$$L(\beta^*, \sigma^2|Y^*) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}(Y^*-X^*\beta^*)^T(Y^*-X^*\beta^*)}$$

(3.5)

Dengan menggunakan metode maximum likelihood, estimasi parameter β^* dan σ^2 dari persamaan (3.5) adalah sebagai berikut:

$$\hat{\beta}^* = (X^{*T}X^*)^{-1}X^{*T}Y^* \quad (3.6)$$

Dan

$$\hat{\sigma}^2 = \frac{1}{n}(Y^* - X^*\beta^*)^T(Y^* - X^*\beta^*) \quad (3.7)$$

Estimator $\hat{\beta}^*$ mempunyai sifat-sifat:

$\hat{\beta}^*$ mempunyai sifat unbias. Bukti:

$$\begin{aligned}\hat{\beta}^* &= (X^{*T}X^*)^{-1}X^{*T}Y^* \\ E(\hat{\beta}^*) &= E((X^{*T}X^*)^{-1}X^{*T}Y^*) \\ &= (X^{*T}X^*)^{-1}X^{*T}E(Y^*) \\ &= (X^{*T}X^*)^{-1}X^{*T}X^*\beta^* \\ &= I\beta^* \\ &= \beta^*\end{aligned}$$

Selanjutnya akan dibuktikan bahwa estimator $\hat{\beta}^*$ adalah estimator efisien. Dikatakan estimator efisien apabila mempunyai nilai variansi yang terkecil. Sehingga $var\left(\hat{\beta}^*\right) = (X^{*T}X^*)^{-1}\sigma^2$ harus sekecil mungkin agar estimator $\hat{\beta}^*$ efisien.

Kemudian sifat estimator yang ketiga yaitu konsisten. Dikatakan estimator yang konsisten jika $lim_{n\to\infty} P\left(\left|\hat{\theta}_n - \theta\right| < \varepsilon = 1\right.$ sehingga:

$$lim_{n\to\infty} E\left(\hat{\theta} - E(\hat{\theta})\right)^2 = \lim_{n\to\infty}(X^{*T}X^*)^{-1}\sigma^2 = 0.$$

Sehingga dapat dikatakan bahwa $\hat{\beta}^*$merupakan estimator yang konsisten

Selanjutnya menentukan Fungsi *likelihood* dari estimator $\hat{\beta}^*$ dan $\hat{\sigma}^2$ adalah sebagai berikut:

$$L(\hat{\beta}^*, \hat{\sigma}^2) = \frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2}(Y^* - X^*\beta^*)^T(Y^* - X^*\beta^*)} \quad (3.8)$$

Fungsi likelihood ini kemudian dilogaritmakan. Sehingga diperoleh:

$$-\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{n}{2} \quad (3.9)$$

3. Pendeteksian *Outlier*

Pendeteksian *outlier* pada regresi nonlinier dengan metode statistik *likelihood displacement* (LD) dilakukan dengan cara menghilangkan pengamatan yang diduga mengandung *outlier* pada model. Misalkan ada k pengamatan yang dikumpulkan dalam suatu himpunan tertentu, dengan k adalah pengamatan yang diduga mengandung *outlier*. Dimana $k < n$. Dan misalkan indeks A_k adalah kumpulan dari k pengamatan yang diduga *outlier* dengan $A_k = i_1, i_2, \dots, i_k$, dan misalkan indeks $i_k \in \{1,2,\dots,n\}$.

Dengan mempertimbangkan pengamatan k dalam estimasi parameter, maka likelihood displacement untuk $\hat{\theta} = \hat{\beta}^*, \hat{\sigma}^2$dan $\hat{\theta}_{[A_k]} = \hat{\beta}^*_{[A_k]}, \hat{\sigma}^2_{[A_k]}$adalah:

$$LD_{A_k}(\beta^*|\sigma^2) = 2\left(\ln L(\hat{\beta}^*, \hat{\sigma}^2) - \ln L(\hat{\beta}^*_{[A_k]}, \hat{\sigma}^2_{[A_k]})\right) \quad (3.10)$$

dimana $\hat{\beta}^*$adalah *maximum likelihood estimation* dari β^* dan $\hat{\sigma}^2$ adalah *maximum likelihood estimation* pada keseluruhan pengamatan dan $\hat{\beta}^*_{[A_k]}$ dan $\hat{\sigma}^2_{[A_k]}$ adalah MLE dari β^* dan σ^2 ketika pengamatan dengan indeks A_k dihilangkan.

Pada kasus khusus yaitu $\theta_1 = (\beta_1^*, \sigma_1^2)$ subset dari $\theta = (\beta^*, \sigma^2)$, maka fungsi *likelihood displacement* dapat dimodifikasi menjadi

$$\begin{aligned}LD_{A_k}&((\beta_1^*, \sigma_1^2)|(\beta_2^*, \sigma_2^2)) \\ &= 2\{ln\ L(\hat{\beta}^*, \hat{\sigma}^2) \\ &\quad - \ln L((\hat{\beta}_1^*, \hat{\sigma}_1^2)_{[A_k]}\ , (\hat{\beta}_2^*, \hat{\sigma}_2^2)(\hat{\beta}_1^*, \hat{\sigma}_1^2)_{[A_k]})\}\end{aligned}$$

Dengan:

$$\begin{aligned}L(&(\hat{\beta}_2^*, \hat{\sigma}_2^2)(\hat{\beta}_1^*, \hat{\sigma}_1^2)_{[A_k]}) \\ &= max_{(\beta_2^*, \sigma_2^2)} L\left((\hat{\beta}_1^*, \hat{\sigma}_1^2)_{[A_k]}\ , (\hat{\beta}_2^*, \hat{\sigma}_2^2)\right)\end{aligned}$$

adalah memaksimumkan fungsi *log likelihood* pada parameter (β_2^*, σ_2^2) dengan $(\beta_1^*, \sigma_1^2) = (\hat{\beta}_1^*, \hat{\sigma}_1^2)_{[A_k]}$ maka $\beta_1^* = \hat{\beta}^*_{1[A_k]}$ dan $\sigma_1^2 = \hat{\sigma}^2_{1[A_k]}$ adalah *maximum likelihood estimation* dari (β_1^*, σ_1^2) ketika pengamatan k dihilangkan.

Selanjutnya untuk keseluruhan data ketika k pengamatan pada himpunan A_k dihilangkan maka modelnya menjadi:

$$Y^*_{[A_k]} = X^*_{[A_k]}\beta^*_{[A_k]} + \varepsilon_{[A_k]} \quad (3.11)$$

dengan $Y^*_{[A_k]} \sim N(0, I\sigma^2)$

estimasi parameter $\beta^*_{[A_k]}$ dan $\sigma^2_{[A_k]}$ dari persamaan (3.11) dengan maximum likelihood diperoleh:

$$\hat{\beta}^*_{[A_k]} = \hat{\beta}^* - (X^{*T}X^*)^{-1}X^{*T}_{A_k}\left(I - Q_{A_k}\right)^{-1}\hat{\varepsilon}^*_{A_k}$$

estimator $\hat{\beta}^*_{[A_k]}$ adalah estimator tak bias. Dan

$$\hat{\sigma}^2_{[A_k]} = \frac{n}{n-k}\sigma^2 + \frac{1}{n-k}\hat{\varepsilon}^T_{A_m}Q_{A_m}\left(I - Q_{A_m}\right)^{-1}\hat{\varepsilon}_{A_k}$$

dengan:

$$Q_{A_k} = X^*_{A_k}(X^{*T}X^*)^{-1}X^{*T}_{A_k}$$

$$\hat{\varepsilon}^*_{A_k} = Y^*_{A_k} - X^*_{A_k}\hat{\beta}^*$$

Pada kasus khusus seperti yang telah dijelaskan maka estimasi dari $\hat{\sigma}^2\left(\hat{\beta}^*{}_{[A_k]}\right)$ dimana $\hat{\sigma}^2\left(\hat{\beta}^*{}_{[A_k]}\right)$ adalah *maximum likelihood estimation* dari $\hat{\sigma}^2$ ketika β^* diestimasi dengan $\hat{\beta}^*{}_{[A_k]}$. Dengan mensubtitusikan $\hat{\beta}^*{}_{[A_k]}$ untuk β^* pada $\hat{\sigma}^2$, sehingga diperoleh:

$$\begin{aligned}\hat{\sigma}^2\left(\hat{\beta}^*{}_{[A_k]}\right) = \hat{\sigma}^2 + \frac{1}{n}\hat{\varepsilon}^{*T}_{A_k}\left(I - Q_{A_k}\right)^{-1}Q_{A_k}\mathrm{x} \\ \left(I - Q_{A_k}\right)^{-1}\hat{\varepsilon}^*_{A_k}\end{aligned}$$

Selanjutnya menentukan Fungsi *likelihood* dari $\hat{\beta}^*{}_{[A_k]}, \hat{\sigma}^2\left(\hat{\beta}^*{}_{[A_k]}\right)$ diperoleh:

$$L\left(\hat{\beta}^*_{[A_k]}, \hat{\sigma}^2\left(\hat{\beta}^*_{[A_k]}\right)\right) = \frac{1}{\left(2\pi^{\frac{n}{2}}\right)\left(\hat{\sigma}^2\left(\hat{\beta}^*_{[A_k]}\right)^{\frac{n}{2}}\right)} e^{-\frac{1}{2\hat{\sigma}^2\left(\hat{\beta}^*_{[A_k]}\right)}\left(Y^*-X^*_{[A_k]}\beta^*_{[A_k]}\right)^T\left(Y^*-X^*_{[A_k]}\beta^*_{[A_k]}\right)}$$

Fungsi likelihood ini kemudian dilogaritmakan. Sehingga diperoleh:

$$-\frac{n}{2} ln\, 2\pi - \frac{n}{2} ln\, \hat{\sigma}^2\left(\hat{\beta}^*_{[A_k]}\right) - \frac{n}{2} \qquad (3.12)$$

4. Metode statistik likelihood displacemen (LD)

Likelihood Displacement dari β^* dan σ^2 yang diberikan pada persamaan (2.5) adalah:

$$LD_{A_k}(\beta^*|\sigma^2) = 2\left\{\ln L(\hat{\beta}^*, \hat{\sigma}^2) - \ln L\left(\hat{\beta}^*_{[A_k]}, \hat{\sigma}^2\left(\hat{\beta}^*_{[A_k]}\right)\right)\right\}$$

Subtitusikan persamaan (3.9) dan (3.12) ke persamaan (2.5) maka:

$$\begin{aligned} LD_{A_k}(\beta^*|\sigma^2) &= 2\left\{\left(-\frac{n}{2} ln(2\pi) - \frac{n}{2} ln(\sigma^2)\right) - \left(-\frac{n}{2} ln\, 2\pi - \frac{n}{2} ln\, \hat{\sigma}^2\left(\hat{\beta}^*_{[A_k]}\right)\right)\right\} \\ &= 2\left\{-\frac{n}{2} ln(2\pi) - \frac{n}{2} ln(\sigma^2) + \frac{n}{2} ln\, 2\pi + \frac{n}{2} ln\, \hat{\sigma}^2\left(\hat{\beta}^*_{[A_k]}\right)\right\} \\ &= 2\left\{-\frac{n}{2} ln(\sigma^2) + \frac{n}{2} ln\, \hat{\sigma}^2\left(\hat{\beta}^*_{[A_k]}\right)\right\} \\ &= -n\, ln\, \sigma^2 + n\ ln\, \hat{\sigma}^2\left(\hat{\beta}^*_{[A_k]}\right) \\ &= n\left\{-\, ln\, \sigma^2 + ln\, \hat{\sigma}^2\left(\hat{\beta}^*_{[A_k]}\right)\right\} \\ &= n\left\{ln\, \hat{\sigma}^2\left(\hat{\beta}^*_{[A_k]}\right) - ln\, \sigma^2\right\} \\ &= n\, ln\left(\frac{\hat{\sigma}^2\left(\hat{\beta}^*_{[A_k]}\right)}{\sigma^2}\right) \\ &= n\, ln\left\{\frac{\hat{\sigma}^2 + \frac{1}{n}\hat{\varepsilon}^{*T}_{A_k}\left(I - Q_{A_k}\right)^{-1} Q_{A_k}\left(I - Q_{A_k}\right)^{-1}\hat{\varepsilon}^*_{A_k}}{\sigma^2}\right\} \end{aligned}$$

misal $C_{A_k} = \left(I - Q_{A_k}\right)^{-1} Q_{A_k}\left(I - Q_{A_k}\right)^{-1}$, maka:

$$\begin{aligned} LD_{A_k} &= n\, ln\left\{\frac{\hat{\sigma}^2 + \frac{1}{n}\hat{\varepsilon}^{*T}_{A_k} C_{A_k}\hat{\varepsilon}^*_{A_k}}{\sigma^2}\right\} \\ &= n\, ln\left\{\frac{\hat{\sigma}^2}{\hat{\sigma}^2} + \frac{\frac{1}{n}\hat{\varepsilon}^{*T}_{A_k} C_{A_k}\hat{\varepsilon}^*_{A_k}}{\hat{\sigma}^2}\right\} \\ &= n\, ln\left\{1 + \frac{1}{n\hat{\sigma}^2}\hat{\varepsilon}^{*T}_{A_k} C_{A_k}\hat{\varepsilon}^*_{A_k}\right\} \end{aligned}$$

Sehingga *Likelihood Displacement* yang diduga mengandung *outlier* adalah sebagai berikut:

$$LD_{A_k} = n\, ln\left\{1 + \frac{1}{n\hat{\sigma}^2}\hat{\varepsilon}^{*T}_{A_k} C_{A_k}\hat{\varepsilon}^*_{A_k}\right\}$$

Untuk menunjukkan keakuratan dari hasil metode LD dalam mendeteksi adanya *outlier*, maka digunakan uji statistik. Uji statistik disini dilakukan dengan cara membandingkan MSE dari metode LD dengan MSE dari regresi pada umumnya (regresi tanpa *outlier*). Statistik uji yang digunakan adalah

$$\Lambda = \frac{MSE_{LD_{A_k}}}{MSE_{reg}} \sim \chi^2_{\lambda_i}$$

dimana λ_i, $i = 1, 2, \ldots, k$, adalah nilai eigen dari C_{A_k}. Ketika nilai $MSE_{LD_{A_k}}$ lebih besar dari pada MSE_{reg} maka nilai Λ akan semakin besar.

Dari hasil uji statistik yang telah dijelaskan, maka diberikan uji hipotesis sebagai berikut:

$H_0: A_k =$ adalah bukan *outlier*

$H_1: A_k =$ adalah *outlier*

H_0 ditolak jika $\chi^2_{hitung} > \chi^2_{tabel}$ dan H_0 diterima jika $\chi^2_{hitung} < \chi^2_{tabel}$.

PENUTUP

Berdasarkan pembahasan yang dipaparkan, dapat disimpulkan bahwa metode statistik *Likelihod Displacement* (LD) mampu mendeteksi adanya *outlier* pada regresi nonlinier multiplikatif.

Sebelum menerapkan metode LD terlebih dahulu harus melinierkan model dengan asumsi bahwa *error* berdisrtibusi normal kemudian mengestimasi parameter regresi nonlinier multiplikatif dengan metode *maximum likelihood estimation*. Kemudian menerapkan metode statistik *likelihood displacement*, sehingga diperoleh hasil perumusan LD_{A_k} *likelihood displacement* untuk pengamatan yang diduga mengandung *outlier*.

Keakuratan metode LD dalam mendeteksi adanya *outlier* ditunjukkan dengan uji statistik. Yaitu dengan membandingkan MSE dari LD dengan MSE dari regresi pada umumnya. Statistik uji yang digunakan adalah Λ. Hipotesis awal ditolak ketika $\chi^2_{hitung} > \chi^2_{tabel}$, Sehingga terbukti LD_{A_k} adalah suatu *outlier*..

DAFTAR PUSTAKA

[1] Abdusysyakir. 2007. *Ketika Kyai Mengajar Matematika*. Malang : UIN-Malang Press.

[2] Al-Asqolani, I. H. & Al-Imam, A. 2007. *Fathul Baari Penjelas Kitab Shahih Al-Bukhari (12).* Penj. Amiruddin. Jakarta: Pustaka Azzam.

[3] Al-Mahally, I. J. & As-Suyuthi, I. J. 1990. *Terjemah Tafsir Jalalain Berikut Asbaabun Nuzul*. Bandung: Sinar Baru.

[4] Al-Maraghi, A. M. 1989. *Tafsir Al-Maraghi*. Semarang: CV. Thoha Putra.

[5] Amrullah, A. A. 1981. *Tafsir Al-Azhar*. Surabaya: Yayasan Latimojong

[6] Draper, N. & Harry, S. 1992. *Analsis Regresi Terapan (edisi kedua)*. Jakarta: PT. Gramedia Pustaka Utama.

[7] Ghoffur, A. dkk. 2007. *Tafsir Ibnu Katsir (8)*. Bogor: Pustaka Imam Syafi'i.

[8] Gujarati, D. N. 2007. *Dasar-dasar Ekonometri jilid 1 edisi ke-3*. Jakarta: Penerbit Erlangga.

[9] Hasan, M. I. 2002. *Pokok-pokok Materi Metodologi Penelitian dan Aplikasinya*. Jakarta:Ghalia Indonesia.

[10] Herrhyanto, N. 2007. *http: // www. Herryanto. blog/ Statistika. Matematika. I. html* (diunduh pada tanggal 26 januari 2012).

[11] Makkulau, S. L. & Purhadi, M. M. 2010. *Pendeteksian Outlier dan Penentuan Faktor-Faktor yang Mempengaruhi Produksi Gula dan Tetes Tebu dengan Metode Likelihood Displacement Statistic-Lagrange.* Jurnal Teknik Industri, Volume 12. No. 2 Desember 2010, 95-100.

[12] Mood, M Alexander dkk.1986. *Introduction to the Theory of Statistics*. McgrawHill Book Company.Sembiring, RK. 1995. *Analisis Regresi.* Bandung: ITB.

[13] Murray & Larry. 2007. *Statistik edisi ke-3*. Jakarta: Erlangga.

[14] Shihab, M. Q. 2003. *Tafsir Al-Mishbah Volume* 14. Jakarta: Lentera Hati.

[15] Sudjana. 2005. *Metoda Statistika*. Bandung: Transito.

[16] Turmudi & Harini, S. 2008. *Metode Statistika Pendekatan Teoritis dan Aplikatif*. Malang: UIN-Press.

[17] Wibisono, Y. 2005. *Metode Statistik*. Yogyakarta: Gadjah Mada University Press.

[18] Xu, A. & Steiner. 1998. Outlier Detection Methods in Multivariate Regression Models. *Journal of Multivariate Analysis*, 65, 1998, pp. 195-208.

[19] Yitnosumarto, S. 1990. *Dasar-Dasar Statistika*. Jakarta: Rajawali.