REGRESI *ROBUST*

Menurut Chen (2002) regresi *robust* adalah salah satu penduga regresi yang *robust* atau resisten dalam menganalisis data yang menyimpang terhadap asumsi analisis regresi. Beberapa penyimpangan terhadap asumsi yang dimaksud misalnya residu yang tidak berdistribusi normal atau adanya pencilan yang mempengaruhi model. Metode ini dibutuhkan karena MKT yang dianggap penduga terbaik dalam analisis regresi ternyata peka terhadap data yang menyimpang dari asumsi.

Prosedur *robust* ditujukan untuk memberikan dugaan yang lebih tepat dan cepat terhadap data yang melanggar asumsi dengan cara meniadakan identifikasi adanya data pencilan, serta dapat menanggulangi data pencilan.

Regresi *robust* merupakan metode regresi yang digunakan ketika distribusi dari nilai residu tidak normal dan adanya beberapa pencilan yang berpengaruh pada model. Metode ini digunakan untuk menganalisa data yang dipengaruhi oleh pencilan sehingga model yang dihasilkan *robust*.

Regresi robust ditujukan untuk mengakomodasi adanya keanehan data, sekaligus meniadakan identifikasi adanya data pencilan dan juga bersifat otomatis dalam menanggulangi data pencilan. Analisis regresi *robust* tidak membuat galat model menjadi normal, tetapi model yang dihasilkan oleh metode *robust* memiliki tingkat keakuratan lebih tinggi.

Regresi *robust* memiliki beberapa estimasi parameter, yaitu estimasi *Maximum Likelihood Type* (M), estimasi *Scale* (S), estimasi *Least Trimmed Square* (LTS), estimasi *Least Median Square* (LMS) dan estimasi *Method of Moment* (MM). Regresi *robust* memiliki beberapa jenis fungsi pembobot, diantaranya adalah Huber dan Tukey Bisquare.

3.1. Estimasi M Huber.

Beberapa peneliti mengembangkan metode untuk mengatasi dampak pencilan apabila metode *OLS* digunakan. Metode ini disebut estimasi-M (1993). Menurut Li, *et al*. (1998) dan Susanti (2014), penggunaan metode kuadrat terkecil tidak akan sesuai dalam menyelesaikan masalah yang mengandung pengamatan outlier atau ekstrim, karena asumsi normalitas tidak dapat dipenuhi. Estimasi-M mengantisipasi hal ini dengan mendefinisikan fungsi ε, $\rho(\varepsilon)$ yang disebut fungsi Huber. Dalam estimasi-M, $b_0, b_1, \ldots, b_k$ masing-masing adalah estimator $\beta_0, \beta_1, \ldots, \beta_k$ dipilih sedemikian rupa sehingga $\sum \rho(\varepsilon)$ minimum dan fungsi Huber didefinisikan sebagai

$$\rho(\varepsilon) = \begin{cases} \varepsilon^2 & \text{jika } -l < \varepsilon < l \\ 2l|\varepsilon| - l^2 & \varepsilon < -l \text{ atau } \varepsilon > l \end{cases}$$

di mana $l = 1{,}5\hat{\sigma}$. Untuk memperkirakan σ digunakan $\hat{\sigma} = 1{,}483\ MAD$ dimana MAD (*Median of Absolute Deviation*) adalah median dari absolut yang tersisa. Untuk mendapatkan nilai estimasi-M, diperlukan algoritma penghitungan.

3.1.1. Algoritma estimasi-M Huber

Algoritma estimasi M berikut diberikan oleh Birkes dan Dodge (1993).

1. Menentukan prediksi model menggunakan metode kuadrat terkecil $\hat{y} = b_0^0 + b_1^0 x_{i1} + \ldots + b_k^0 x_{ik}$, dan menghitung $e_i^0 = y_i - \hat{y}_i^0$ dan kemudian menghitung $\hat{\sigma} = 1{,}483\ MAD$.
2. Memotong e_i^0 di mana $e_i^* = 1{,}5\ \hat{\sigma}$ jika $e_i^0 > 1{,}5\ \hat{\sigma}$ dan $e_i^* = -1{,}5\ \hat{\sigma}$ jika $e_i^0 < -1{,}5\ \hat{\sigma}$.
3. Menghitung $y_i^* = e_i^* + \hat{y}_i^0$ dan kemudian temukan nilai b^0 dengan menggunakan metode kuadrat terkecil. Proses iterasi berlanjut sampai nilai yang diperoleh b^0 sama dengan iterasi sebelumnya. Nilai b^0 diperoleh dari regresi dengan estimasi-M.

Uji Signifikansi Model *Reduce*

Hipotesis :

$H_0: \beta_j = 0\ \forall_j, j = q+1, q+2, \ldots, k$

$H_1: \exists_j \in \beta_j \neq 0\ \forall_j, j = q+1, q+2, \ldots, k$

dengan q merupakan jumlah variabel independen beta *i* yang dimasukkan dalam model, k merupakan jumlah variabel independen yang digunakan. Sehingga model menjadi

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_q X_{iq} + \varepsilon_i$$

Statistik uji:

$$F_m = \frac{STR_{reduce} - STR_{full}}{(k-q)\hat{\lambda}}$$

$$\hat{\lambda} = \frac{(\frac{n}{m}) \sum (e_i *)^2}{n-k-1}$$

dengan *STR (Sum of Transformed Residuals)* merupakan jumalh yang tersisa ditransformasikan, dimana STR_{reduce} dan STR_{full} diperoleh dari model *full* dan model *reduce*. Algoritma *STR full*:

1. Menghitung $e_{i,}$ dan $\hat{\sigma} = 1.483\ MAD$
2. Memotong nilai e_i untuk mendapatkan nilai e_i dari $\rho(\varepsilon)$

   dengan

   $$\rho(\varepsilon) = \begin{cases} \varepsilon^2 & \text{jika } -l < \varepsilon < l \\ 2l|\varepsilon| - l^2 & \text{jika } \varepsilon < -l \text{ atau } \varepsilon > l \end{cases}$$

Selanjutya, menghitung *STR reduce* dengan cara sama dengan menghitung *STR full, m* merupakan jumlah e_i yang tidak dipotong. Ho ditolak jika Fm > F tabel dengan derajat bebas *(k-q);(n-k-1)*

3.2.Estimasi S

Estimasi S pertama kali diperkenalkan oleh Rousseeuw & Yohai (1984). Estimasi S merupakan estimasi *robust* yang dapat mencapai *breakdown point* yang tinggi yaitu 50%. *Breakdown point* adalah ukuran umum proporsi dari pencilan yang dapat ditangani sebelum pengamatan tersebut mempengaruhi model. Estimasi S dapat mencapai *breakdown point* hingga 50% maka estimasi S dapat mengatasi setengah dari pencilan dan memberikan pengaruh yang baik bagi pengamatan lainnya.

Estimasi S didefinisikan sebagai $\hat{\beta}_s = \hat{\sigma}_s(e_1, e_2, \cdots e_n)$. $\hat{\beta}$ adalah estimator koefisien regresi, $e_1, e_2, \cdots e_n$ merupakan sesatan, dan nilai estimator skala *robust* yang dinotasikan dengan $\hat{\sigma}_s$ yang minimum dan memenuhi

$$min \sum_{i=1}^{n} \rho \left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta}{\hat{\sigma}_s} \right) \tag{3.2}$$

Estimator parameter skala *robust* pada estimasi S dinyatakan sebagai

$$\hat{\sigma}_s = \begin{cases} \dfrac{median\ |e_i - median\ (e_i)|}{0.6745} & ,iterasi = 1 \\ \sqrt{\dfrac{1}{nK} \sum_{i=1}^{n} w_i\, {e_i}^2} & ,iterasi > 1 \end{cases} \tag{3.3}$$

dimana $K = 0.199$. Fungsi ρ merupakan fungsi objektif *Tukey bisquare*

$$\rho(u_i) = \begin{cases} \frac{u_i^{\,2}}{2} - \frac{u_i^{\,4}}{2c^2} + \frac{u_i^{\,6}}{6c^4} & , |u_i| \leq c \\ \frac{c^2}{6} & , |u_i| > c \end{cases}$$

dengan

$$u_i = \frac{e_i}{\hat{\sigma}_s} \tag{3.4}$$

Penyelesaian persamaan (3.2) adalah dengan cara menurunkannya terhadap β sehingga menghasilkan persamaan

$$\sum_{i=1}^{n} \rho'\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta}{\hat{\sigma}_s}\right) = 0 \qquad j = 0,1,\dots,k$$

$$\sum_{i=1}^{n} x_{ij}\psi\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta}{\hat{\sigma}_s}\right) = 0 \qquad j = 0,1,\dots,k$$

ψ disebut fungsi pengaruh yang merupakan turunan dari ρ. Sehingga bisa dituliskan $\rho' = \psi$ yaitu

$$\psi(u_i) = \rho'(u_i) = \begin{cases} \frac{2u_i}{2} - \frac{4u_i^{\,3}}{2c^2} + \frac{6u_i^{\,5}}{6c^4} & , |u_i| \leq c \\ 0 & , |u_i| > c \end{cases}$$

$$= \begin{cases} u_i - \frac{2u_i^{\,3}}{c^2} + \frac{u_i^{\,5}}{c^4} & , |u_i| \leq c \\ 0 & , |u_i| > c \end{cases}$$

$$= \begin{cases} u_i\left(1 - \frac{2u_i^{\,2}}{c^2} + \frac{u_i^{\,4}}{c^4}\right) & , |u_i| \leq c \\ 0 & , |u_i| > c \end{cases}$$

$$= \begin{cases} u_i\left(1 - \frac{u_i^{\,2}}{c^2}\right)^2 & , |u_i| \leq c \\ 0 & , |u_i| > c \end{cases}$$

$$= \begin{cases} u_i\left(1 - \left(\frac{u_i}{c}\right)^2\right)^2 & , |u_i| \leq c \\ 0 & , |u_i| > c \end{cases}$$

dengan w_i merupakan fungsi pembobot *Iteratively Reweighted Least Square* (IRLS) dan $c = 1.547$

$$w_i(u_i) = \frac{\psi(u_i)}{u_i} = \begin{cases} \frac{u_i\left(1-\left(\frac{u_i}{c}\right)^2\right)^2}{u_i} & , |u_i| \le c \\ 0 & , |u_i| > c \end{cases}$$

$$= \begin{cases} \begin{cases} \left[1-\left(\frac{u_i}{c}\right)^2\right]^2 & , |u_i| \le c \\ 0 & , |u_i| > c \end{cases} & , iterasi = 1 \\ \frac{\rho(u_i)}{{u_i}^2} & , iterasi > 1 \end{cases} \tag{3.5}$$

Pembobot Huber dan Pembobot Tukey Bisquare secara ringkas disajikan pada Tabel 3.3. berikut :

Tabel 3.3. Fungsi objektif dan fungsi pembobot untuk Huber dan Tukey Bisquare

Metode	Fungsi Objektif	Fungsi Pembobot	Interval
Huber	$\rho(u_i) = \begin{cases} \frac{1}{2} u_i^2 \\ c\,\lvert u_i\rvert - \frac{c^2}{2} \end{cases}$	$w(u_i) = \begin{cases} 1 \\ \frac{c}{\lvert u_i\rvert} \end{cases}$	$\lvert u_i\rvert \le c$ $\lvert u_i\rvert > c$
Tukey Bisquare	$\rho(u_i) = \begin{cases} \frac{{u_i}^2}{2} - \frac{{u_i}^4}{2c^2} + \frac{{u_i}^6}{6c^4} \\ \frac{c^2}{6}, \end{cases}$	$w(u_i) = \begin{cases} \left[1-\left(\frac{u_i}{c}\right)^2\right]^2 \\ 0 \end{cases}$	$\lvert u_i\rvert \le c$ $\lvert u_i\rvert > c$

3.2.1. Algoritma Estimasi S

Algoritma yang dilakukan dalam mengestimasi parameter model regresi dengan estimasi S

1. Menentukan estimasi awal dengan MKT sehingga diperoleh nilai awal $\hat{\beta}^0$
2. Menghitung nilai sesatan $e_i = y_i - \hat{y}_i$
3. Menghitung nilai $\hat{\sigma}_s$ dengan rumus pada persamaan (3.3)
4. Menghitung u_i dengan rumus pada persamaan (3.4)
5. Menghitung nilai fungsi pembobot Tukey bisquare dengan rumus pada persamaan (3.5)
6. Menghitung $\hat{\beta}_s$ menggunakan metode IRLS dengan pembobot w_i sehingga diperoleh sesatan yang baru

7. Menjadikan sesatan langkah (6) sebagai sesatan awal langkah (2) sehingga diperoleh nilai u_i dan w_i yang baru. Iterasi diulang sampai mendapatkan nilai yang konvergen dari $\hat{\beta}_s$

Metode Estimasi *Least Trimmed Square* (LTS).

Least Trimmed Square (LTS) diusulkan oleh Rousseuw (1984) sebagai alternatif *robust* untuk mengatasi kelemahan MKT, yaitu dengan menggunakan sebanyak h ($h \leq n$) kuadrat sesatan yang diurutkan nilainya (Rousseuw, 1984).

$$\min_{b} \sum_{i=1}^{h} e_i^2$$

dengan $h = \left[\frac{n}{2}\right] + \left[\frac{k+2}{2}\right]$

e_i^2= kuadrat sesatan yang diurutkan dari terkecil ke terbesar.

n = banyaknya sampel.

k = banyaknya variable independen.

Pada proses perhitungan, nilai h harus selalu dalam bentuk bilangan bulat. Oleh karena itu, jika nilai h bukan dalam bentuk bilangan bulat maka dilakukan pembulatan ke atas

Menurut Rousseeuw dan Hubert (1997) estimator untuk LTS dapat didapatkan dengan cara

$$s_{LTS} = d_{h,n}\sqrt{\frac{1}{h}\sum_{i=1}^{h} e_i^2}$$

dengan

$$d_{h,n} = 1/\sqrt{1 - \frac{2n}{hc_{h,n}}\phi\left(\frac{1}{c_{h,n}}\right)}$$

$$c_{h,n} = 1/\Phi^{-1}\left(\frac{h+n}{2n}\right)$$

Setelah didapatkan nilai s_{LTS} lalu menghitung pembobot w_i dengan rumus

$$w_i = \begin{cases} 0 & \text{jika } \left|\frac{e_i}{s_{LTS}}\right| > 2{,}5 \\ 1 & \text{untuk lainnya.} \end{cases}$$

3.3.1. Algoritma Estimasi LTS

Algoritma yang harus dilakukan dalam estimasi LTS adalah:

1. Mengestimasi koefisien regresi $\widehat{\beta_i}$ dengan metode kuadrat terkecil dengan i = 1,2,… n.
2. Menghitung n kuadrat sesatan e_i^2 yang bersesuaian dengan $\widehat{\beta_i}$.
3. Menghitung sejumlah $h = \left[\frac{n}{2}\right] + \left[\frac{k+2}{2}\right]$ dengan i = 1,2,… n.
4. Melakukan estimasi parameter $\hat{\beta}_{baru(i)}$ dari $h_{baru(i)}$ pengamatan.
5. Menghitung $\hat{\beta}_{baru(i)}$ dan ulangi menghitung h berikutnya sampai konvergen.
6. Menghitung pembobot w_i.

Mengestimasi koefisien regresi $\hat{\beta}_{LTS}$ dengan pembobot w_i.

.Metode Estimasi *Least Median of Square* (LMS)

Prinsip dasar metode regresi robust estimasi LMS adalah mencocokkan sebagian besar data setelah pencilan teridentifikasi sebagai titik yang tidak berhubungan dengan data (Rousseeuw & Leroy, 1987). Jika pada MKT hal yang perlu dilakukan adalah meminimumkan kuadrat residual ($\sum_{i=1}^{n} e_i^2$), maka pada LMS hal yang perlu dilakukan adalah meminimumkan median kuadrat sesatan yaitu

$$Mj = \{ e_i^2 \} = \{M_1, M_2, \dots , Ms \}$$

dengan e_i^2 adalah kuadrat sesatan hasil estimator dengan MKT.

Cara untuk mendapatkan nilai M_1 adalah dengan mencari himpunan bagian dari matriks X sejumlah hi pengamatan, yaitu

$$h_i = h_1 = \left[\frac{n}{2}\right] + \left[\frac{k+2}{2}\right]$$

di mana n banyaknya data, dan k banyaknya variable independen.

Pada proses perhitungan, nilai h_i harus selalu dalam bentuk bilangan bulat. Oleh karena itu, jika nilai hi bukan dalam bentuk bilangan bulat maka dilakukan pembulatan ke atas. Cara untuk mencari M_2, ditentukan himpunan bagian data dari matriks X sejumlah h_2 pengamatan, yaitu

$$h_i = h_1 = \left[\frac{n}{2}\right] + \left[\frac{k+2}{2}\right] \text{ di mana } n = h_1 .$$

LMS merupakan penduga pada regresi *robust*, maka sama hal nya dengan penduga lain pada regresi *robust*, prinsip dasar dari LMS adalah dengan memberikan bobot wii pada data sehingga data pencilan tidak mempengaruhi model parameter taksiran. (Rousseuw, 1984)

Berdasarkan Rousseeuw (1984) , bobot wii dirumuskan dengan ketentuan sebagai berikut

$$w_{ii} = \frac{\psi(\varepsilon_i^*)}{\varepsilon_i^*}$$

Pembobot w_{ii} ditentukan berdasarkan fungsi dari pembobot yaitu

$$\psi(\varepsilon_i^*) = \begin{cases} \varepsilon_i^* \text{ jika } |\varepsilon_i^*| \le 2.5 \\ 2.5 \text{ jika } \varepsilon_i^* > 2.5 \\ -2.5 \text{ jika } \varepsilon_i^* < -2.5 \end{cases}$$

dengan $\varepsilon_i^* = \frac{e_i}{\hat{\sigma}}$ dan $\hat{\sigma} = 1.4826\left[1 + \frac{5}{n-p}\right]\sqrt{M_j}$

setelah bobot wii dihitung, dapat dibentuk matriks W sebagai berikut

$$W = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{pmatrix}$$

dengan entri matriks wij = 0 , dimana $i \neq j$.

Setelah terbentuk matriks W, maka penaksir parameter regresi LMS dapat dihitung dengan menggunakan rumus

$$\widehat{\boldsymbol{\beta}}_{\mathbf{LMS}} = (\mathbf{X^T W X})^{-1}(\mathbf{X^T W Y})$$

3.4.1. Algoritma estimasi LMS

Algoritma dalam estimasi LMS adalah :

1. Mengestimasi koefisien regresi $\widehat{\beta_1}$ dengan metode kuadrat terkecil.
2. Menghitung nilai kuadrat residu (e^2).
3. Menghitung median dari kuadrat residu.
4. Menghitung nilai $h = \left[\frac{n}{2}\right] + \left[\frac{k+2}{2}\right]$.
5. Mengurutkan variabel berdasarkan kuadrat sesatan terkecil, lalu potong sesuai nilai h.
6. Melakukan estimasi β_{baru} dengan h pengamatan.
7. Melakukan langkah pertama hingga keenam sampai didapatkan nilai β yang konvergen.
8. Menghitung pembobot w_{ii}.
9. Mengestimasi koefisien regresi $\hat{\beta}_{LMS}$ dengan pembobot $w_{ii.}$

.Estimasi MM .

Estimasi-MM merupakan gabungan gabungan dari metode estimasi-S dan metode estimasi-M. Langkah pertama dalam estimasi-MM adalah mencari estimator-S kemudian menetapkan parameter-parameter regresi menggunakan estimasi M. Estimator *robust* pada prinsipnya merupakan estimasi yang meminimumkan suatu fungsi sesatan ρ (fungsi objektif). Fungsi ρ merupakan representasi pembobot dari sesatan, dapat ditulis :

$$\hat{\beta}_{MM} = arg\ \min_{\beta} \sum_{i=1}^{n} \rho(e_i) = arg \min_{\beta} \sum_{i=1}^{n} \rho(y_i - \sum_{j=0}^{k} x_{ij}\beta_i)$$

Estimator ini biasanya diperoleh berdasarkan estimator skalanya, σ, yang akan diestimasi secara bersama-sama, seperti :

$$\hat{\beta}_{M} = arg\ \min_{\beta} \sum_{i=1}^{n} \rho(u_i) = arg \min_{\beta} \sum_{i=1}^{n} \rho\left(\frac{\varepsilon_i}{\hat{\sigma}}\right)$$

$$\hat{\beta}_{M} = arg\ \min_{\beta} \sum_{i=1}^{n} \rho\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta_j}{\hat{\sigma}}\right)$$

Suatu estimasi skala *robust* memperlukan estimasi, yaitu :

$$\hat{\sigma} = \frac{MAD}{0{,}6745} = \frac{\text{median } |e_i - \text{median}(e_i)|}{0{,}6745}$$

Fungsi pembobot pada regresi *robust* estimasi-MM yang digunakan adalah fungsi pembobot Tukey Bisquare. Kriteria fungsi pembobot Tukey Bisquare, didapat :

$$\rho'(u_i) = \begin{cases} u_i\left(1 - \left(\frac{u_i^2}{c^2}\right)^2\right)^2 & , |u_i| \le c \\ 0 & , |u_i| > c \end{cases}$$

1.5.1. Algoritma Estimasi MM

Adapun lalgoritma yang dilakukan dalamestimasi MM adalah :

1) Estimasi parameter model dengan metode kuadrat terkecil.
2) Pengujian asumsi regresi linear berganda.
3) Pendeteksi pencilan dengan Metode *DFITS*.
4) Metode regresi *robust* dengan estimasi MM :
   a. Menghitung estimasi parameter $\hat{\beta}_{jS}$ dengan regresi *robust* estimasi S, langkah-langkahnya sebagai berikut :
      i. Menghitung nilai e_i.
      ii. Mencari nilai $\hat{\sigma}$.
      iii. Menghitung nilai u_i.

iv. Mencari nilai pembobot $w(u_i)$dengan menggunakan kriteria fungsi pembobot Tukey Bisquare.

v. Mencari nilai parameter $\hat{\beta}_S$ regresi robust metode WLS (*weight least square*) dengan fungsi pembobot w_i.

vi. Mengulangi langkah i sampai iv hingga didapatkan $\hat{\beta}_{jS}$ yang konvergen.

b. Menghitung estimasi parameter $\hat{\beta}_{jMM}$ dengan regresi *robust* estimasi M, langkah-langkahnya sebagai berikut :

i. Menghitung nilai e_i.

ii. Menghitung nilai $\hat{\sigma}$.

iii. Menghitung nilai u_i.

iv. Menghitung nilai pembobot $w(u_i)$menggunakan kriteria fungsi pembobot Tukey Bisquare.

v. Mencari nilai parameter $\hat{\beta}_{MM}$ menggunakan regresi *robust* metode *weight least square* (WLS) dengan fungsi pembobot w_i.

vi. Mengulangi langkah ke-i sampai ke-iv hingga didapatkan nilai $\hat{\beta}_{jMM}$ yang konvergen