PERBANDINGAN METODE ESTIMASI-*M*, ESTIMASI-*S*, DAN ESTIMASI-*MM* PADA MODEL REGRESI *ROBUST* UNTUK MEMPREDIKSI PRODUKSI KEDELAI DI INDONESIA

Jurnal

Diajukan kepada Fakultas Matematika dan Ilmu Pengetahuan Alam

Universitas Negeri Yogyakarta

untuk Memenuhi Sebagian Persyaratan

guna Memperoleh Gelar Sarjana Sains

Oleh

Wening Dyah Pratitis

NIM 12305144023

PROGRAM STUDI MATEMATIKA

JURUSAN PENDIDIKAN MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS NEGERI YOGYAKARTA

2016

PERSETUJUAN

Jurnal dengan judul

PERBANDINGAN METODE ESTIMASI-*M*, ESTIMASI-*S*, DAN ESTIMASI-*MM* PADA MODEL REGRESI *ROBUST* UNTUK MEMPREDIKSI PRODUKSI KEDELAI DI INDONESIA

Yang disusun oleh,

Nama	:	Wening Dyah Pratitis
NIM	:	12305144023
Prodi	:	Matematika

Telah disetujui Dosen Pembimbing dan direview oleh Dosen Penguji untuk

Memenuhi sebagian persyaratan guna memperoleh Gelar Sarjana Sains

Yogyakarta, 25 Juli 2016

Direview Dosen Penguji	Disetujui Dosen Pembimbing
Dr. Dhoriva U.W NIP.196603311993032001	**Dra. Endang Listyani, M.S** NIP.197206221998022001

PERBANDINGAN METODE ESTIMASI-M, ESTIMASI-S, DAN ESTIMASI-MM PADA MODEL REGRESI *ROBUST*

COMPARISON OF M-ESTIMATION, S-ESTIMATION, AND MM-ESTIMATION METHODS OF ROBUST REGRESSION

Oleh: Wening Dyah Pratitis[1)], Endang Listyani[2)]
Program Studi Matematika, Jurusan Pendidikan Matematika, FMIPA UNY
[1)]wendyp31@yahoo.co.id [2)]listy_matuny@yahoo.com

Abstrak

Analisis regresi linier adalah analisis terhadap hubungan satu variabel dependen (Y) dengan satu atau lebih variabel independen (X). Estimasi parameter yang sering digunakan adalah Metode Kuadrat Terkecil (MKT). MKT memberikan nilai *error* yang besar apabila data mengandung pencilan. untuk itu diperlukannya model regresi *robust* yang dapat menangani keberadaan pencilan. Permasalahan yang akan dikaji adalah membandingkan metode estimasi-M, estimasi-S, dan estimasi-MM pada model regresi *robust*. Tujuan dari penelitian ini adalah untuk membandingkan manakah dari ketiga metode estimasi yang paling efektif untuk memprediksi produksi kedelai di Indonesia. Dalam penelitian ini diambil data produksi kedelai sebagai variabel Y dan variabel independen yang meliputi luas panen, suhu, curah hujan, kelembaban, dan lama penyinaran. Keefektifan suatu metode dapat dibandingkan dengan melihat nilai R^2. Dari hasil pembahasan diperoleh metode estimasi paling efektif untuk memprediksi produksi kedelai di Indonesia adalah metode estimasi-S dengan nilai R^2 sebesar 97,73%, estimasi-MM sebesar 82,22%, dan estimasi-M sebesar 67,38%.

Kata kunci: analisis regresi, metode kuadrat terkecil, regresi *robust*, estimasi-M, estimasi-S, estimasi-MM.

Abstract

Linear regression analysis is an analysis of the relationship of one dependent variable (Y) with one or more independent variables (X). Frequently used parameter estimation is the Ordinary Least Squares (OLS). OLS is not precise when the data has outlier, it needs robust regression models that can handle the presence of outlier. The issue that will be examined is to compare methods of M-estimation, S- estimation and MM-estimation of robust regression model. The purpose of this study is to compare which of these three methods of estimation is the most effective way to predict the production of soybean in Indonesia. In this study, The Y variable is the soybean production data, and the X variables include crop land area, temperature, precipitation, humidity, and length shines. The effectiveness of a method can be compared by looking at the value of R^2. From the results, it is obtained that the most effective estimation methods to predict the production of soybean in Indonesia is S-estimation method with R^2 value of 97,73%, MM-estimation of 82,22%, and M-estimation of 67,38%.

Keywords: regression analysis, ordinary least square, robust regression, M-estimation, S-estimation, MM-estimation

PENDAHULUAN

Analisis regresi merupakan metode analisis yang dapat digunakan untuk menganalisis data dan mengambil kesimpulan yang bermakna tentang hubungan ketergantungan variabel satu terhadap variabel lainnya Salah satu metode estimasi dalam regresi linier yang sering digunakan adalah metode kuadrat terkecil (MKT) atau sering disebut *Ordinary Least Squares* (OLS). Penggunaan metode ini membutuhkan beberapa asumsi yang harus dipenuhi untuk menghasilkan model yang baik, yaitu model estimasi linier tidak bias terbaik atau *Best Linier Unbiased Estimator* (BLUE), yaitu tidak terdapat multikolinieritas, homoskedastisitas, dan tidak terdapat autokorelasi. Masalah yang sering terjadi pada data amatan adalah adanya pencilan (*outlier*). Keberadaan pencilan pada suatu data dapat mengganggu proses analisis data, sehingga pendeteksian pencilan merupakan hal yang sangat penting untuk dilakukan.

Adanya pencilan dalam data dapat mengakibatkan estimasi koefisien regresi yang diperoleh kurang efisien karena memberikan nilai *error* yang besar apabila menggunakan MKT. Sedangkan tindakan membuang begitu saja suatu pencilan bukanlah tindakan yang tepat karena ada kalanya pencilan memberikan informasi yang cukup berarti. Oleh karena itu, diperlukan suatu estimasi yang lebih efisien dalam menangani suatu pencilan yang dikenal dengan regresi *robust*. Regresi *robust* merupakan metode regresi yang digunakan ketika distribusi dari sisaan tidak normal dan/atau adanya beberapa pencilan yang berpengaruh pada model (Ryan, 1997:150).

Regresi *robust* adalah metode yang penting untuk menganalisis data yang mengandung pencilan. Untuk menaksir parameter regresi *robust* dapat digunakan beberapa metode antara lain: S-estimator, M-estimator, MM-estimator, *Least Median of Squares* (LMS), dan *Least Trimmed Squares* (LTS) (Chen, 2002:1).

Pada penelitian ini, akan dibahas regresi *robust* dengan estimasi-M, estimasi-S, dan estimasi-MM, serta ditunjukkan estimasi yang paling efektif dengan membandingkan nilai koefisien determinasi dari ketiga metode tersebut untuk memodelkan data pada produksi kedelai di Indonesia. Tujuan dari penelitian ini adalah untuk membandingkan keefektifan model melalui nilai R^2.

Regresi Linier

Regresi linier merupakan suatu metode analisis statistik yang mempelajari pola hubungan antara dua variabel atau lebih menggunakan model persamaan linier, sehingga salah satu variabel pada model regresi dapat diduga dari variabel lainnya.

Model regresi linier sederhana ini merupakan suatu model regresi dasar yang melibatkan satu variabel independen saja. Bentuk umum regresi linier sederhana dapat dituliskan sebagai berikut:

$$y_i = \beta_0 + \beta_1 x_i + e_i \qquad (1)$$

(Draper & Smith, 1998:22) dengan y_i merupakan variabel dependen pada observasi ke-i, x_i adalah konstanta yang diketahui yaitu nilai variabel independen yang diketahui, β_0 dan β_1 adalah parameter koefisien regresi, sedangkan e_i merupakan suatu *error*.

Menurut Montgomery & Peck (1992:53), Model regresi linier berganda dengan k variabel independen adalah sebagai berikut:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + e_i$$

atau dapat ditulis

$$y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + e_i \qquad (2)$$

untuk $i = 1,2,\dots,n$

dengan:

y_i = nilai variabel dependen pada observasi ke-i

$\beta_0, \beta_1, \dots, \beta_k$ = parameter koefisien regresi

x_{ij} = nilai variabel independen yang ke-j pada observasi ke-i

e_i = *random error*

Uji Asumsi

Pada penelitian ini, diambil simulasi data dari Badan Pusat Statistika (BPS) dan Dinas Pertanian. Data yang diambil adalah data dari 34 propinsi di Indonesia meliputi produksi kedelai sebagai variabel dependen, dan beberapa variabel independen yang meliputi luas area panen, suhu, curah hujan, kelembaban, dan lama penyinaran.

Pada model regresi, perlu dilakukan uji asumsi analisis regresi untuk mengetahui apakah model baik (memenuhi asumsi) atau tidak. Asumsi yang memenuhi analisis regresi dengan MKT antara lain: data berdistribusi normal, homoskedastisitas, non autokorelasi, dan non multikolinieritas.

Identifikasi *Outlier*

Untuk mengidentifikasi adanya *outlier*, penulis menggunakan pengujian residu *Jackknife* atau *externally studendized residual* atau *R-student.* Menurut Chatterjee & Hadi (1986:380), definisi *Jackknife* atau biasa juga disebut sebagai *R-student*, yang dilambangkan dengan t_i adalah:

$$t_i = e_i(\widehat{\sigma}_i) = \frac{e_i}{\widehat{\sigma}_{(i)}\sqrt{1-h_{ii}}} \qquad (4)$$

dengan $p = variabel + 1$ dan t_i berdistribusi t_{n-p-1} jika model asumsi terpenuhi dan $e_i \sim N(0, \sigma^2 I)$. Notasi h_{ii} merupakan elemen diagonal ke-i dari matriks *hat* dan e_i merupakan residu ke-i.

Metode Kuadrat Terkecil

Setelah pengujian asumsi dan *outlier*, langkah selanjutnya yaitu membuat persamaan regresi dengan metode kuadrat terkecil.

Parameter $\beta_0, \beta_1, \dots, \beta_k$ tidak diketahui dan perlu ditentukan nilai estimasinya. Menurut Montgomery & Peck (1992:112), Metode Kuadrat Terkecil (MKT) digunakan untuk mengestimasi koefisien $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ yaitu dengan meminimumkan jumlah kuadrat galat. Fungsi yang meminimumkan adalah:

$$S(\beta_0, \beta_1, \dots, \beta_k) = \sum_{i=1}^{n} {e_i}^2$$

$$= \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{k}\beta_j x_{ij}\right)^2 \qquad (3)$$

Fungsi S akan diminimalkan dengan menentukan turunannya terhadap $\beta_0, \beta_1, \dots, \beta_k$, harus memenuhi

$$\frac{\partial S}{\partial \beta} = -2 \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{k}\beta_j x_{ij}\right) = 0$$

Notasi matriks yang diberikan pada Persamaan (2) adalah

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{e}$$

dengan

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}; \qquad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix};$$

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}; \qquad e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

Regresi *Robust*

Regresi *robust* merupakan metode regresi yang digunakan ketika distribusi dari sisaan tidak normal dan/atau adanya beberapa pencilan yang berpengaruh pada model (Ryan, 1997:150). Metode ini merupakan alat penting untuk menganalisa data yang dipengaruhi oleh pencilan sehingga dihasilkan model yang *robust* atau *resistance* terhadap pencilan.

Estimasi-*M*

Menurut Montgomery, Peck, & Vining (2006:372), Estimasi-M adalah suatu kelas dari regresi *robust* yang meminimalkan suatu fungsi residu ρ sebagai berikut:

$$\widehat{\beta}_M = \min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \rho(e_i)$$

$$\widehat{\beta}_M = \min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \rho\left(y_i - \sum_{j=0}^{k} x_{ij}\beta_j \right) \quad (5)$$

Fungsi ρ merupakan fungsi representasi pembobot dari residu. fungsi ini merupakan gabungan dari MKT dan *Least Absolute Value* (LAV). Fungsi Huber lebih resisten terhadap *outlier* daripada MKT. Sehingga fungsi ρ yang digunakan adalah fungsi objektif Huber dengan persamaan sebagai berikut:

$$\rho(u_i) = \begin{cases} \frac{1}{2}(u_i)^2 & , |u_i| \le c \\ c|u_i| - \frac{1}{2}c^2 & , |u_i| > c \end{cases} \quad (6)$$

dengan $\rho(u_i)$ adalah fungsi simetris dari residu atau fungsi yang memberikan kontribusi pada masing-masing residu pada fungsi objektif, dengan $u_i = \frac{e_i}{\hat{\sigma}_s}$.

Untuk meminimalkan Persamaan (5), akan digunakan turunan parsial pertama fungsi ρ terhadap $\beta_j (j = 0, 1, \ldots, k)$ sama dengan 0, sehingga menghasilkan suatu kondisi minimum. Sehingga diperoleh

$$\sum_{i=1}^{n} x_{ij}\psi\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta_j}{\hat{\sigma}_s} \right) = 0 \quad (7)$$

dengan $\psi = \rho'$ dan x_{ij} merupakan observasi ke-i pada *regressor* ke-j dan $x_{i0} = 1$. Berdasarkan definisi, fungsi pembobot $w(u_i)$ dari persamaan (7), yaitu:

$$w(u_i) = \frac{\psi\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta_j}{\hat{\sigma}_s} \right)}{\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta_j}{\hat{\sigma}_s} \right)} \quad (8)$$

karena nilai $u_i = \frac{e_i}{\hat{\sigma}_s}$ sebagai pengganti e_i, maka persamaan (8) menjadi:

$$w_i = w(u_i) = \frac{\psi(u_i)}{(u_i)} = \begin{cases} 1 & , |u_i| \le c \\ \frac{c}{|u_i|} & , |u_i| > c \end{cases}$$

untuk fungsi pembobot Huber, konstanta yang digunakan adalah $c = 1{,}345$ (Fox, 2002:3). Pada umumnya, fungsi ψ merupakan fungsi yang tidak linear dan Persamaan (7) harus diselesaikan dengan menggunakan *iteratively reweighted least squares* (IRLS), maka diperoleh

$$\sum_{i=1}^{n} x_{ij}\psi\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta_j}{\hat{\sigma}_s} \right) = 0 \quad (9)$$

sehingga

$$\sum_{i=1}^{n} x_{ij} w_{i0}\left(y_i - \sum_{j=0}^{k} x_{ij}\beta_j \right) = 0, \quad (10)$$

dengan

$$w_{i0} = \begin{cases} 1 & , y_i = \sum_{j=0}^{k} x_{ij}\beta_j \\ \dfrac{\psi\left[\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta_j}{\hat{\sigma}_s} \right]}{\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta_j}{\hat{\sigma}_s}} & , y_i \ne \sum_{j=0}^{k} x_{ij}\beta_j \end{cases} \quad (11)$$

Untuk kasus regresi berganda, persamaan (10) menjadi:

$$\sum_{i=1}^{n} x_{ij} w_{i0} \left(y_i - \sum_{j=0}^{k} x_{ij} \beta_j \right) = 0$$

$$\sum_{i=1}^{n} x_{ij} w_{i0} y_i - \sum_{i=1}^{n} \sum_{j=0}^{k} x_{ij}^2 w_{i0} \beta_j = 0$$

$$\sum_{i=1}^{n} \sum_{j=0}^{k} x_{ij}^2 w_{i0} \beta_j = \sum_{i=1}^{n} x_{ij} w_{i0} y_i \qquad (12)$$

dalam notasi matriks, persamaan (12) di atas menjadi:

$$\boldsymbol{X^T W_0 X \beta} = \boldsymbol{X^T W_0 Y} \qquad (13)$$

dengan $\boldsymbol{W}_0$ adalah matriks bobot berordo $n \times n$ dengan elemen diagonal utama $w_{10}, w_{20}, \ldots, w_{n0}$ yang diberikan pada Persamaan (11). Persamaan (13) dikenal sebagai persamaan normal *weighted least squares* (WLS). Oleh karena itu, estimator satu-langkah (*one-step estimator*) adalah:

$$\boldsymbol{\beta} = (\boldsymbol{X^T W}_0 \boldsymbol{X})^{-1} \boldsymbol{X^T W}_0 \boldsymbol{Y} \qquad (14)$$

Langkah berikutnya adalah menghitung ulang bobot dari Persamaan (3.9) tetapi menggunakan $\widehat{\boldsymbol{\beta}}_1$ bukan $\widehat{\boldsymbol{\beta}}_0$. Pada umumnya, untuk $\boldsymbol{W}_j$ bobot yang diberikan dapat menyelesaikan:

$$\widehat{\boldsymbol{\beta}}_{j+1} = \left(\boldsymbol{X^T W}_j \boldsymbol{X}\right)^{-1} \boldsymbol{X^T W}_j \boldsymbol{Y} \qquad (15)$$

Iterasi terus dilakukan hingga mencapai konvergen, yaitu hingga selisih nilai $\widehat{\boldsymbol{\beta}}_{j+1}$ dengan $\widehat{\boldsymbol{\beta}}_j$ mendekati 0.

Estimasi-*S* (*Scale*)

Menurut Montgomery, Peck, & Vining, (2006:388), estimasi-S akan meminimumkan suatu ukuran skala residu, dapat didefinisikan hasil estimasi dari persamaan berikut:

$$\widehat{\beta}_s = \min_\beta \widehat{\sigma}_s \left[e_1(\beta), e_2(\beta), \ldots, e_n(\beta)\right] \qquad (16)$$

dengan e_i merupakan residu ke-i dari β dan $\widehat{\sigma}_s(e_1, e_2, \ldots, e_n)$ didefinisikan sebagai solusi dari:

$$\frac{1}{n} \sum_{i=1}^{n} \rho\left(\frac{e_i}{\hat{\sigma}_s}\right) = \frac{1}{n} \sum_{i=1}^{n} \rho\left(\frac{y_i - \sum_{j=0}^{k} x_{ij} \beta_j}{\hat{\sigma}_s}\right);$$

$$\frac{1}{n} \sum_{i=1}^{n} \rho\left(\frac{y_i - \sum_{j=0}^{k} x_{ij} \beta_j}{\hat{\sigma}_s}\right) = K \qquad (17)$$

dengan K merupakan suatu konstanta yang didefinisikan sebagai K =0,1995 (Alma, 2011:415). Persamaan (17) dibagi dengan $\left(\frac{e_i}{\widehat{\sigma}_s}\right)^2$ sehingga menghasilkan

$$\hat{\sigma}_s = \sqrt{\frac{1}{nK} \sum_{i=1}^{n} w_i e_i^2} \qquad (18)$$

Salah satu fungsi ρ (pembobot) untuk Persamaan (17) adalah fungsi *Tukey bisquare* (Rousseeuw & Leroy, 1987). Berikut merupakan fungsi pembobot *Tukey bisquare:*

$$\rho(u_i) = \begin{cases} \dfrac{u_i^2}{2} - \dfrac{u_i^4}{2c^2} + \dfrac{u_i^6}{6c^4} & untuk\ |u_i| \le c \\ \dfrac{c^2}{6} & untuk\ |u_i| > c \end{cases}$$

dengan $\rho(u_i)$ adalah fungsi simetris dari residu yaitu fungsi yang memberikan kontribusi pada masing-masing residu pada fungsi objektif. Turunan dari fungsi ρ adalah:

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

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

sehingga diperoleh w_i yang merupakan suatu fungsi pembobot *iteratively reweighted least squares* (IRLS)

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

dengan $u_i = \frac{e_i}{c}$ dan agar diperoleh *breakdown point* secara konsisten mendektai 50%, maka ditetapkan c = 1,547 (Fox, 2002:3). Persamaan (17) pada estimasi-S dapat diselesaikan dengan metode IRLS hingga mencapai konvergen.

Estimasi-MM

Metode estimasi ini dikenalkan oleh Yohai (1987) yang menggabungkan suatu *high breakdown point* (50%) dengan efisiensi tinggi (95%).

Alur dari estimasi-MM dapat diuraikan sebagai berikut:

1) Mengestimasi koefisien $\hat{\beta}_j^{(1)}$, sehingga diperoleh residual $e_i^{(1)}$ yang diambil dari regresi *robust* dengan *high breakdown point*.
2) Residual $e_i^{(1)}$ pada langkah pertama digunakan untuk menghitung skala residual estimasi-M, $\hat{\sigma}$ dan dihitung pula bobot awal $w_i^{(1)}$.
3) Residual $e_i^{(1)}$ dan skala residual $\hat{\sigma}$ dari langkah (2) digunakan dalam iterasi awal dengan metode WLS untuk menghitung koefisien regresi.

$$\sum_{i-1}^{n} w_i^{(1)}\left(\frac{e_i^{(1)}}{\hat{\sigma}}\right)x_i = 0$$

dengan w_i menggunakan pembobot Huber atau Tukey Bisquare
4) Menghitung bobot baru $w_i^{(2)}$ menggunakan residual dari iterasi awal WLS (langkah 3).
5) Mengulang langkah 2, 3, 4 diulang (reiterasi dengan skala residual tetap konstan) sampai $\sum_{i=1}^{n}\left|e_i^{(m)}\right|$ konvergen, yaitu selisih nilai $\hat{\beta}^{(m+1)}$ dengan $\hat{\beta}_j^{(m)}$ mendekati 0, dengan m adalah banyaknya iterasi.

Estimasi-MM didefinisikan sebagai berikut

$$\hat{\beta}_{MM} = \underbrace{\arg min}_{\beta} \sum_{i=1}^{n} \rho\left(\frac{e_i}{\hat{\sigma}_s}\right)$$

$$\hat{\beta}_{MM} = \underbrace{\arg min}_{\beta} \sum_{i=1}^{n} \rho\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta_j}{\hat{\sigma}_s}\right) \quad (19)$$

Untuk meminimunkan Persamaan (19), turunan parsial pertama fungsi ρ terhadap $\beta_j (j = 0,1, \dots, k)$ akan sama dengan 0, sehingga menghasilkan suatu kondisi perlu untuk minimum. Dengan mendefinisikan fungsi pembobot yang sama, Persamaan (19) dapat ditulis sebagai berikut

$$\sum_{i=1}^{n} x_{ij} w_i \left(y_i - \sum_{j=0}^{k} x_{ij}\beta_j\right) = 0 \qquad (20)$$

Selanjutnya, Persamaan (20) akan diselesaikan menggunakan IRLS. Estimasi awal koefisien $\hat{\beta}^{(1)}$ dan $e_i^{(1)}$, diambil dari regresi *robust* dengan *high breakdown point* (estimasi-S). Untuk pembobot permulaan $w_i^{(1)} = w\left(e_i^{(1)}\right)$, maka Persamaan (20) dapat ditulis sebagai berikut:

$$\sum_{i=1}^{n} x_{ij} w_i^{(1)} \left(y_i - \sum_{j=0}^{k} x_{ij}\beta_j\right) = 0,$$

$$j = 0,1, \dots, k \qquad (21)$$

dengan

$$w_i^{(1)} = \begin{cases} \dfrac{\psi\left(\dfrac{y_i - \sum_{j=0}^{k} x_{ij}\hat{\beta}_j^{(1)}}{\hat{\sigma}_s}\right)}{\dfrac{y_i - \sum_{j=0}^{k} x_{ij}\hat{\beta}_j^{(1)}}{\hat{\sigma}_s}} & , jika\ y_i \neq \sum_{j=0}^{k} x_{ij}\hat{\beta}_j^{(1)} \\ 1 \quad , jika\ y_i = \sum_{j=0}^{k} x_{ij}\hat{\beta}_j^{(1)} & \end{cases}$$

Untuk kasus regresi berganda, Persamaan (20) menjadi:

$$\sum_{i=1}^{n}\sum_{j=0}^{k} x_{ij}{}^{2} w_i{}^{(1)} \hat{\beta}_j = \sum_{i=1}^{n} w_i{}^{(1)} x_i y_i$$

dalam notasi matriks, Persamaan (20) di atas menjadi:

$$\boldsymbol{X'W^{(1)}X\beta} = \boldsymbol{X'W^{(1)}Y} \qquad (21)$$

maka, estimator satu langkah dapat ditulis sebagai berikut:

$$\widehat{\boldsymbol{\beta}}^{(2)} = (\boldsymbol{X'W^{(1)}X})^{-1}\boldsymbol{X'W^{(1)}Y} \qquad (22)$$

Pada langkah selanjutnya, dihitung kembali bobot dari $w_i{}^{(2)}$ menggunakan $\hat{\beta}_j{}^{(2)}$ dan skala residual $\hat{\sigma}_s$. Untuk $w_i{}^{(m)}$ bobot yang diberikan, dapat diperoleh estimator $\hat{\beta}^{(m+1)} = (\boldsymbol{X'W^{(m)}X})^{-1}\boldsymbol{X'W^{(m)}Y}$. Biasanya dibutuhkan beberapa iterasi hingga $\sum_{i=1}^{n}\left|e_i{}^{(m)}\right|$ konvergen, yaitu selisih nilai $\hat{\beta}^{(m+1)}$ dengan $\hat{\beta}_j{}^{(m)}$ mendekati 0.

HASIL EMPIRIK

Pada penelitian ini, diambil data dari Badan Pusat Statistik dan Dinas Pertanian. Produksi kedelai sebagai variabel independen (Y), dan variabel dependen (X) meliputi luas area panen (X_1), suhu (X_2), curah hujan (X_3), kelembaban (X_4), dan lama penyinaran (X_5).

Langkah pengolahan data diawali dengan uji asumsi, pemodelan dengan MKT, uji outlier, dan menentukan model estimasi-M, estimasi-S, dan estimasi-MM serta membandingkan nilai efektifitas antara ketiga metode tersebut.

Uji Asumsi

1. Uji Normalitas

Berdasar hasil *output* SPSS, uji *Kolmogorov-Sminov* pada data yang diperoleh nilai signifikansi (Asymp. Sig. 2-Tailed) sebesar 0,118 lebih besar dari 0.05 = α, maka residu berdistibusi normal.

2. Uji Homoskedastisitas

Uji homoskedastisitas dalam hal ini menggunakan *rank spearman*. Berdasar hasil output SPSS, diperoleh sebagai berikut:

Tabel 2. Hasil Uji Homoskedastisitas

Variabel Independen	***Sig (2-Tailed) Unstandardized Residual***
X_1 (Luas area panen)	0,059 > 0,05
X_2 (Suhu)	0,066 > 0,05
X_3 (Curah Hujan)	0,345 > 0,05
X_4 (Kelembaban)	0,258 > 0,05
X_5 (Lama Penyinaran)	0,153 > 0,05

Pada Tabel 2. diperoleh nilai *sig* pada masing-masing variabel independen lebih dari α = 0,05, maka masing-masing variabel tidak terjadi heteroskedastisitas.

3. Uji Non Autokorelasi

Berdasar hasil output SPSS, uji *Durbin-Watson* menghasilkan nilai d sebesar 2,096, dengan dU = 1,8076 dan $(4 - dU)$ = 2,1924. Artinya nilai d berada diantara dU dan $(4 - dU)$ sehingga dapat disimpulkan bahwa model regresi tidak terdapat autokorelasi.

4. Uji Non Multikolinieritas

Uji non multikolinieritas dilakukan dengan membandingkan nilai *tolerance* dan VIF. Berdasar hasil *output* SPSS, diperoleh hasil uji non multikolinieritas sebagai berikut:

Tabel 3. Hasil Uji Non Multikolinieritas

Variabel Independen	***Tolerance***	**VIF**
X_1 (Luas area panen)	0,947 > 0,1	1,056 < 10

X_2 (Suhu)	0,235 > 0,1	4,252 < 10
X_3 (Curah Hujan)	0,676 > 0,1	1,479 < 10
X_4 (Kelembaban)	0,365 > 0,1	2,740 < 10
X_5 (Lama Penyinaran)	0,387 > 0,1	2,582 < 10

Pada Tabel 3 diperoleh masing-masing nilai tolerance lebih besar dari 0,1 dan nilai VIF kurang dari 10, sehingga dapat diartikan bahwa tidak adanya multikolinieritas antara variabel independen.

Metode Kudrat Terkecil

Untuk memudahkan perhitungan, penulis menggunakan program SAS 9.1, diperoleh model regresi yang dapat dibentuk sebagai berikut:

$$\hat{Y}_{MKT} = -2136{,}63382 + 1{,}65022X_1 - 67{,}55636X_2 - 3{,}78511X_3 + 69{,}48093X_4 - 483{,}48411X_5$$

dengan:

$\hat{Y}$ = Produksi kedelai

X_1= Luas area panen

X_2= Suhu

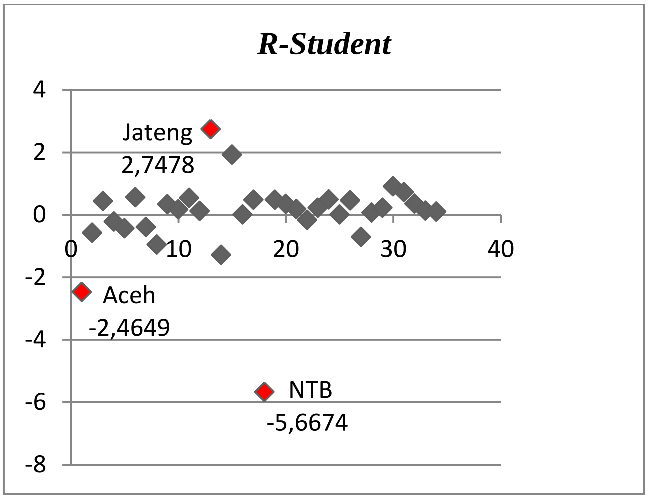
X_3= Curah hujan

X_4= Kelembaban

X_5= Lama Penyinaran

Uji Outlier

Dengan taraf nyata α = 5%, p = 6 dan n = 34, maka dari tabel distribusi t telah diperoleh $t_{\frac{\alpha}{2};(n-p-1)} = 2{,}5018$, sehingga nilai didapatkan $-2{,}0518 < |R - student| < 2{,}0518$, dengan menggunakan bantuan *software*, diperoleh nilai *R-Student* sebagai berikut



Gambar 1. Plot Nilai Residu *Jackknife* untuk Setiap Observasi

Dari Gambar 1 dapat dilihat bahwa nilai *R-student* pada observasi ke$-i$ (t_i), yaitu observasi pertama yaitu Provinsi Aceh (t_1), Jawa Tengah sebagai observasi ke-13 (t_{13}) dan Nusa Tenggara Barat sebagai obsetvasi ke-18 (t_{18}) terletak pada daerah penolakan, maka observasi ke-1, ke-13 dan ke-18 merupakan *outlier*. Dengan demikian, metode kuadrat terkecil tidak dapat digunakan untuk melihat pengaruh dari variabel independennya. Oleh karena itu, metode estimasi-M, estimasi-S, dan estimasi-MM yang akan digunakan untuk menyelesaikan kasus ini. Ketiga metode tersebut akan dibandingkan satu dengan yang lainnya.

Estimasi-*M*

Dengan menggunakan program SAS 9.1, telah diperoleh model regresi sebagai berikut:

$$\hat{Y}_M = -1901{,}65 + 1{,}6596X_1 - 211{,}516X_2 - 5{,}0494X_3 + 101{,}1419X_4 - 144{,}070X_5$$

Estimasi-*S*

Dengan menggunakan bantuan *software* SAS 9.1 untuk menghitung estimasi dalam regresi *robust*, model regresi yang dapat dibentuk adalah sebagai berikut:

$$\hat{Y}_S = -960{,}84 + 1{,}4424X_1 - 106{,}18X_2 - 3{,}4728X_3 + 58{,}5809X_4 - 129{,}65X_5$$

Estimasi-MM

Dengan menggunakan program SAS 9.1, telah diperoleh model regresi sebagai berikut:

$$\hat{Y}_{MM} = -967{,}961 + 1{,}4452X_1 - 108{,}804X_2 - 3{,}4777X_3 + 58{,}7784X_4 - 116{,}122X_5.$$

Hasil Perbandingan Estimasi Parameter Regresi *Robust* Estimasi-M, Estimasi-S, dan Estimasi-MM pada Data Produksi Kedelai di Indonesia Tahun 2014

Setelah mengetahui model regresi dari masing-masing parameter, selanjutnya akan dibandingkan hasil dari ketiga metode estimasi yaitu antara metode estimasi-M, estimasi-S dan estimasi-MM. Dengan demikian dapat diketahui metode mana yang memberikan hasil terbaik terhadap model.

Tabel 8. Hasil Perbandingan Estimasi Parameter Regresi *Robust* antara Estimasi-M, Estimasi-S, dan Estimasi-MM pada Data Produksi Kedelai di Indonesia Tahun 2014

Metode	*R-Square*
Estimasi-M	67,38
Estimasi-S	97,73
Estimasi-MM	82,22

Berdasarkan Tabel 8 di atas dapat dilihat bahwa nilai *R-square* (R^2), dapat disimpulkan bahwa nilai $R-square_S > R-square_{MM} > R-square_M$ yaitu $97{,}73\% > 82{,}22\% > 67{,}38\%$ yang berarti bahwa variansi yang dapat dijelaskan oleh model hasil estimasi-MM yaitu 82,22% lebih tinggi dari estimasi-M 67,38%, dan nilai R^2 hasil estimasi-S sebesar 97,73% paling tinggi di antara model kedua estimasi pembandingnya yaitu estimasi-M dan estimasi-MM.

SIMPULAN DAN SARAN

Simpulan

Dari pembahasan yang telah dipaparkan penulis, dapat disimpulkan bahwa berdasarkan ketiga metode estimasi, telah diperoleh nilai *R-square* dari estimasi-M, estimasi-S, dan estimasi-MM, dengan nilai $R-square_S > R-square_{MM} > R-square_M$ yaitu $97{,}73\% > 82{,}22\% > 67{,}38\%$. Dengan demikian dapat disimpulkan bahwa regresi *robust* dengan metode estimasi-S merupakan penaksir yang memberikan model paling efektif dari ketiga estimasi tersebut terhadap data faktor-faktor yang mempengaruhi produksi kedelai di Indonesia pada tahun 2014.

Saran

1. Metode estimasi-M, estimasi-S, dan estimasi-MM merupakan metode yang baik dalam menangani masalah pencilan dengan memberikan nilai koefisien determinasi yang tinggi, namun perlu dipelajari juga metode estimasi yang lain dan pengaplikasiannya dalam menangani pencilan, seperti metode LMS dan LTS.
2. Untuk penelitian selanjutnya, sekiranya dapat menggunakan teknik data influence serta bisa lebih memperhatikan faktor-faktor yang mempengaruhi produksi kedelai di Indonesia sehingga dapat menghasilkan model regresi yang lebih baik lagi.

DAFTAR PUSTAKA

Alma, O. G. (2011). Comparison of Robust Regression Methods in Linear Regression. *Int. J. Contemp. Math. Sciences*, 409-421.

Badan Pusat Statistik. (2014). *Produksi Kedelai di Indonesia*. www.bps.go.id. Diakses pada tanggal 28 Januari 2016 pukul 20.00 WIB.

Chatterjee, S., & Hadi, A.S. (1986). Influential Observations, High Leverage Points, and Outliers in Linear Regression. *Statistical Science*, Vol. 1 (3), 379-393.

Chen, C., (2002). *Robust Regression and Detection with The Robustreg Procedure*. Sugi Paper 265-27, SAS Institute, Cary NC, www2.sas.com.

Dinas Tanaman Pangan dan Holtikultura. (2016). *Mungkinkah Mewujudkan Swasembada Kedelai?*. www.dishantor.rokanhulukab.go.id. Diakses pada tanggal 27 Januari 2016 pukul 20.56 WIB.

Draper, N.R., & Smith, H. (1998). *Applied Regression Analysis*. New York: John Wiley and sons.

Fox, J. (2002). “Robust Regression. Apendix to An R and S-Plus Companion to Applied Regression”. January, 2002.

Hanna Ardiyanti. (2011). Perbandingan Keefektifan Metode Regresi *Robust* Estimasi-M dan Estimasi-MM karena Pengaruh *Outlier* dalam Analisis Regresi Linier. *Skripsi*. Universitas Negeri Semarang.

Montgomery, D.C. & Peck, E.A. (1992). *Introduction to Linear Regression Analysis*. Toronto: John Wiley & Sons.

Montgomery, D.C., Peck, E.A., & Vining, G.G. (2006). *Introduction to Linear Regression Analysis. 4th Ed.* Toronto: John Wiley & Sons.

Rousseeuw, P.J., & Leroy, A.M. (1987). *Robust Regression and Outlier Detection.* New York: John Wiley & Sons.

Ryan, T.P. (1997). Modern *Regression Analysis for Scientists and Engineers*. Ghaitersburg: NIST.

Yuliana Susanti., Hasih Pratiwi., & Sri Sulistijowati H. (2013). Optimasi Model Regresi Robust Untuk Memprediksi Produksi Kedelai Di Indonesia. *Prosiding Seminar Nasional Matematika dan Pendidikan Matematika, Yogyakarta 9 November 2013*, M253-M262.