ISBN : 978 – 979 – 16353 – 9 – 4

S - 31

OPTIMASI MODEL REGRESI *ROBUST* UNTUK MEMPREDIKSI PRODUKSI KEDELAI DI INDONESIA

Yuliana Susanti [1], Hasih Pratiwi [2], Sri Sulistijowati H.[3]
[1,2,3]Jurusan Matematika FMIPA, Universitas Sebelas Maret, Surakarta
[1]yuliana.susanti@ymail.com, [2]hasihpratiwi@ymail.com, [3]ssulistijowati@yahoo.com

Abstrak

Regresi *robust* adalah bentuk analisis regresi yang dirancang untuk menghindari beberapa keterbatasan metode parametrik dan non-parametrik tradisional. Metode kuadrat terkecil untuk model regresi sangat sensitif terhadap pencilan. Hal ini biasanya tidak masalah jika pencilan berasal dari sebuah pengamatan ekstrim dari ekor distribusi normal, tetapi jika pencilan berasal dari kesalahan pengukuran atau pelanggaran asumsi maka sangat mempengaruhi keabsahan hasil regresi. Dalam analisis regresi, adanya produksi kedelai yang jauh melampaui produksi secara umum dapat dikategorikan sebagai pencilan, sehingga penggunaan metode kuadrat terkecil untuk mengestimasi parameter regresi kurang tepat. Untuk mengatasi hal ini diperlukan metode estimasi parameter yang bersifat *robust*. *Robust* diartikan sebagai ketidaksensitifan atau ketegaran terhadap perubahan-perubahan kecil dari asumsi. Artikel ini memberikan gambaran metode estimasi-M, estimasi-S, dan estimasi-MM dalam regresi *robust*, menjelaskan langkah-langkah estimasi parameter, dan menerapkan metode-metode tersebut untuk menentukan model regresi produksi kedelai di Indonesia.

Kata kunci: regresi *robust*, estimasi-M, estimasi-S, estimasi-MM

A. PENDAHULUAN

Kedelai adalah salah satu komoditas penting dalam sembilan kebutuhan pokok. Kedelai bagi industri pengolahan pangan di Indonesia banyak digunakan sebagai bahan baku pembuatan tahu, tempe, kecap, dan susu. Untuk memenuhi kebutuhan masyarakat terhadap tahu dan tempe tersebut, pada saat ini terdapat 115.000 pengrajin tahu dan tempe di seluruh Indonesia berdasarkan data Sensus Ekonomi Nasional (Susenas) 2006 oleh Badan Pusat Statistik (BPS). Produksi kedelai pada tahun 2007 (593 juta ton) sudah berada di bawah produksi kedelai pada tahun 2003 (672 juta ton) atau turun sebesar 79 juta ton. Namun untuk tahun 2009, produksi kedelai kembali naik menjadi 975 juta ton, karena berbagai terobosan yang dilakukan pemerintah, seperti perluasan areal tanam, pemberian bantuan benih maupun sarana produksi pertanian serta insentif bagi petani agar mereka bergairah menanam kedelai. Akan tetapi menurut BPS (Aram II), untuk tahun 2010 produksi kedelai kembali turun yaitu 927 juta ton. Penurunan produksi ini disebabkan oleh penurunan luas panen sebesar 45 ribu ha, tetapi di tahun 2010 produktivitas mengalami kenaikan sebesar 0,19 kuintal per ha.

Kebijakan harga pangan merupakan salah satu instrumen penting dalam menciptakan ketahanan pangan nasional. Mengingat pentingnya upaya pemenuhan kebutuhan pangan, khususnya kedelai, diperlukan usaha untuk mengetahui ketersediaan kedelai pada tahun yang akan datang. Kebijakan harga makanan pokok merupakan salah satu instrumen penting dalam menciptakan ketahanan pangan nasional. Mengingat pentingnya upaya pemenuhan kebutuhan pangan, diperlukan usaha untuk memprediksi produksi pada tahun yang akan datang.

Ada beberapa metode yang dapat digunakan untuk memprediksi produksi kedelai serta menyelidiki faktor-faktor yang mempengaruhinya, diantaranya adalah analisis regresi. Analisis regresi merupakan suatu teknik statistik yang digunakan untuk menyelidiki dan memodelkan hubungan antara variabel bebas dan variabel tak bebas. Jika Y variabel tak bebas dan $X_1, X_2, \ldots, X_k$ variabel-variabel bebas, maka model regresi linear secara umum dapat dinyatakan sebagai $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon$ dengan $\beta_0, \beta_1, \ldots,$ danβ_k adalah parameter-parameter regresi dan ε adalah eror yang berdistribusi normal dengan mean nol dan variansi sama. Permasalahan yang muncul dalam analisis regresi adalah menentukan estimator terbaik untuk $\beta_0, \beta_1, \ldots,$ dan$\beta_{k,}$ yang sangat dipengaruhi oleh penggunaan metode. Sebagai contoh, penggunaan metode kuadrat terkecil (MKT) tidak akan tepat dalam menyelesaikan permasalahan yang mengandung observasi pencilan atau ekstrem, karena asumsi kenormalan tidak dapat dipenuhi. Dalam analisis regresi, adanya produksi kedelai yang jauh melampaui produksi secara umum dapat dikategorikan sebagai pencilan, sehingga penggunaan metode kuadrat terkecil untuk mengestimasi parameter regresi kurang tepat. Untuk mengatasi hal ini diperlukan metode estimasi parameter yang bersifat *robust*. *Robust* diartikan sebagai ketidaksensitifan atau ketegaran terhadap perubahan-perubahan kecil dari asumsi. Estimasi dengan metode *maximum likelihood estimator* (MLE) akan menghasilkan estimator yang bersifat sama seperti metode kuadrat terkecil, artinya MLE juga tidak *robust* terhadap pengaruh pencilan. Suatu teknik *robust* yang sering digunakan adalah estimasi-M, S, dan MM. Estimasi *robust* ini merupakan perluasan dari MLE. Tujuan penelitian ini adalah menentukan model regresi optimal untuk memprediksi produksi kedelai di Indonesia dengan menggunakan estimasi-M, estimasi-S, dan estimasi-MM.

B. REGRESI ROBUST

Regresi *robust* merupakan metode regresi yang digunakan ketika distribusi dari sisaan tidak normal atau ada beberapa pencilan yang berpengaruh pada model.Metode ini merupakan alat penting untuk menganalisis data yang dipengaruhi oleh pencilan sehingga dihasilkan model yang kekar terhadap pencilan (Draper and Smith, 1998).Ketika peneliti menyusun model regresi dan melakukan uji asumsi sering ditemui bahwa asumsi regresi dilanggar, transformasi yang dilakukan tidak akan menghilangkan atau melemahkan pengaruh dari pencilan yang akhirnya prediksi menjadi bias. Dalam keadaan ini, regresi *robust* yang tahan terhadap pengaruh pencilan adalah metode yang terbaik. Regresi *robust* digunakan untuk mendeteksi pencilan dan memberikan hasil yang resisten terhadap adanya pencilan (Chen, 2002).

Estimasi-M

Dalam regresi *robust* salah satu metode estimasi yang terkenal adalah estimasi-M. Huruf M menunjukkan bahwa estimasi-M adalah estimasi 'tipe maksimum *likelihood*'. Estimasi-M memenuhi sifat sebagai estimator tak bias dan memiliki variansi minimum dalam kumpulan estimator. Jadi estimator M memiliki variansi terkecil dibandingkan dengan variansi estimator yang lain. Jika estimator pada estimasi-M adalah $\tilde{\beta} = \beta_n\left(x_1,\ldots,x_n\right)$ maka

$$E\left[\beta_n\left(x_1,\ldots,x_n\right)\right] = \beta\,. \tag{1}$$

Persamaan (1) menunjukkan bahwa estimator $\tilde{\beta} = \beta_n\left(x_1,\ldots,x_n\right)$ pada estimasi-M bersifat tak bias. Variansinya merupakan variansi terkecil dibandingkan dengan variansi estimator yang lain yaitu

$$Var\ \left(\hat{\hat{\beta}}\right) \geq \frac{\left[\tilde{\beta}\ '\right]^2}{nE\left(\frac{d}{d\beta} ln\ \ f\left(x_i;\beta\right)\right)^2} \tag{2}$$

dengan $\hat{\hat{\beta}}$ merupakan estimator alternatif yang linear dan tak bias bagi β. Estimasi-M merupakan perluasan dari MLE dan merupakan estimasi yang *robust* (Yuliana dan Susanti, 2008). Pada metode ini dimungkinkan untuk mengeliminasi beberapa data, akan tetapi dalam beberapa kasus tidak selalu tepat dilakukan apalagi jika yang dieliminasi tersebut merupakan data penting atau bibit unggul, yang kasusnya sering ditemui dalam bidang pertanian (Susanti, dkk., 2009; Susanti dan Pratiwi, 2012). Menurut Montgomery dan Peck (2006), pada prinsipnya estimasi-M merupakan estimasi yang meminimumkan suatu fungsisisaan ρ

$$\hat{\beta}_M = min_\beta \sum_{i=1}^{n} \rho(e_i) = min_\beta \sum_{i=1}^{n} \rho\left(y_i - \sum_{j=0}^{k} x_{ij}\,\beta_j\right) \quad (3)$$

Untuk memperoleh persamaan (3), dengan menyelesaikan persamaan

$$\underset{\beta}{min} \sum_{i=1}^{n} \rho(u_i) = min \sum_{i=1}^{n} \rho\left(\frac{e_i}{\sigma}\right) = \underset{\beta}{min} \sum_{i=1}^{n} \rho\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta_j}{\sigma}\right) \quad (4)$$

dengan dipilih estimasi untuk σ adalah

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

Pemilihan konstanta 0,6745 membuat $\hat{\sigma}$ suatu estimator yang mendekati tak bias dari σ jika n besar dan sisaan berdistribusi normal (Montgomery dan Peck, 2006). Fungsi ρ yang digunakan adalah fungsi objektif Tukey *bisquare*

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

Untuk meminimumkan persamaan (3), dicari turunan parsial pertama dari $\hat{\beta}_M$terhadapβ sehingga diperoleh persamaan

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

$$\sum_{i=1}^{n} x_{ij}\psi\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta}{\hat{\sigma}}\right) = 0, \qquad j = 0, 1, \dots, k \quad (5)$$

dengan $\psi = \rho'$ dan x_{ij} adalah observasi ke-i pada variabel bebas ke-j dan $x_{i0} = 1$.

Draper dan Smith (1998) memberikan penyelesaikan persamaan (5), yaitu dengan mendefinisikan suatu fungsi pembobot

$$w(e_i) = \frac{\psi\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta}{\hat{\sigma}}\right)}{\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta}{\hat{\sigma}}\right)}. \quad (6)$$

Karena nilai $u_i = \frac{e_i}{\hat{\sigma}}$ sebagai pengganti e_i, maka persamaan (6) menjadi

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

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

Untuk fungsi pembobot *Tukey bisquare*, konstanta yang digunakan adalah $c = 4{,}685$. Dengan demikian persamaan (5) menjadi

$$\sum_{i=1}^{n} x_{ij}w_i\left(y_i - \sum_{j=0}^{k} x_{ij}\beta\right) = 0, \qquad j = 0, 1, \dots, k. \quad (7)$$

Persamaan (7) dapat diselesaikan dengan metode MKT terboboti secara iterasi yang dinamakan *IterativelyReweighted Least Squares* (IRLS).Untuk menggunakan IRLS, diasumsikan bahwa suatu estimasi awal, $\hat{\beta}^0$ ada dan $\hat{\sigma}_i$ suatu estimasi skala. Untuk j parameter, dengan j adalah jumlah parameter yang akan diestimasi, maka persamaan (7) menjadi

$$\sum_{i=1}^{n} x_{ij} w_i^{0} \left(y_i - \sum_{j=0}^{k} x_{ij}\beta^0\right) = 0, \quad j = 0, 1, \ldots, k. \tag{8}$$

Dalam notasi matriks, persamaan (8) dapat ditulis menjadi

$$X'W_iX\beta = X'W_iY \tag{9}$$

dengan W_i adalah matriks berukuran $n \times n$ dengan elemen-elemen diagonal yang berisi pembobot. Persamaan (9) dikenal sebagai persamaan *Weighted Least Squares (WLS)*. Penyelesaian persamaan tersebut akan memberikan estimator untuk $\hat{\beta}$ yaitu $\hat{\beta} = (X'W_iX)^{-1}(X'W_iY)$. Estimator-M untuk $\hat{\beta}$diperoleh dengan cara melakukan iterasi sampai diperoleh suatu nilai yang konvergen.

Berikut ini merupakan algoritma penghitungan nilai estimasi-M.

1. Melakukan estimasi koefisien regresi pada data menggunakan MKT.
2. Menguji asumsi klasik dari model regresi.
3. Mendeteksi adanya pencilan dalam data.
4. Mengestimasi koefisien regresi *robust* menggunakan estimasi-M.
   a. Menghitung parameter $\hat{\beta}^0$ dengan MKT.
   b. Menghitung nilai sisaan $e_i = y_i - \hat{y}_i$.
   c. Menghitung nilai$\hat{\sigma}_i = \frac{MAD}{0,6745} = \frac{median|e_i - median(e_i)|}{0,6745}$
   d. Menghitung nilai $u_i = \frac{e_i}{\hat{\sigma}_i}$.
   e. Menghitung pembobot
   $$w_i = \begin{cases} \left[1 - \left(\frac{u_i}{4,685}\right)^2\right]^2 & , |u_i| \le 4,685 \\ 0 & , |u_i| > 4,685 \end{cases}.$$
   f. Menghitung parameter $\hat{\beta}_M$ dengan metode *Weighted Least Squares* (*WLS*) dengan pembobot w_i .
   g. Mengulangi langkah b - f sampai diperoleh nilai $\hat{\beta}_M$ yang konvergen.
   h. Uji hipotesis untuk mengetahui apakah variabel bebas mempunyai pengaruh yang signifikan terhadap variabel tak bebas.

Estimasi-S

Selain estimasi-M, dalam regresi *robust* dikenal juga estimasi-S. Estimasi-S pertama kali diperkenalkan oleh Rousseeuw dan Yohai(1984), dandinamakan estimasi-S karena estimasi ini berdasarkan pada skala sisaan dari estimasi-M. Estimasi-Sdidefinisikan sebagai $\hat{\beta}_s = \min_{\beta} \hat{\sigma}_s(e_1, e_2, \ldots, e_n)$ dengan menentukan nilai estimator skala *robust* $(\hat{\sigma}_s)$ yang minimum dan memenuhi

$$min \sum_{i=1}^{n} \rho\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta}{\hat{\sigma}_s}\right) \tag{10}$$

dengan

$$\hat{\sigma}_s = \sqrt{\frac{1}{nK}\sum_{i=1}^{n} w_i\, e_i^2}$$

K= 0, 199, $w_i = w_{\sigma}(u_i) = \frac{\rho(u_i)}{u_i^2}$, dan dipilih estimasi awal

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

Penyelesaian persamaan (10)adalah dengan cara mencari turunannya terhadap βsehingga diperoleh

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

$$\sum_{i=1}^{n} x_{ij}\psi\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta}{\hat{\sigma}_s}\right) = 0, \qquad j = 0, 1, \dots, k \qquad (11)$$

ψ disebut fungsi pengaruh yang merupakan turunan dariρ $(\rho' = \psi)$, turunan dari fungsi ρ adalah

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

dengan w_i merupakan fungsi pembobot IRLS

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

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

dengan $u_i = \frac{e_i}{\hat{\sigma}_s}$danc = 1,547. Persamaan (11) dapat diselesaikan dengan IRLS sehingga mencapai konvergen.

Adapun algoritma penghitungan nilai estimasi-S sebagai berikut:

1. Melakukan estimasi koefisien regresi pada data menggunakan MKT.
2. Menguji asumsi klasik dari model regresi.
3. Mendeteksi adanya pencilan dalam data.
4. Mengestimasi koefisien regresi *robust* menggunakan estimasi-S.
   a. Menghitung parameter $\hat{\beta}^0$ dengan MKT.
   b. Menghitung nilai sisaan $e_i = y_i - \hat{y}_i$.
   c. Menghitung nilai

$$\hat{\sigma}_i = \begin{cases} \frac{median|e_i - median(e_i)|}{0{,}6745} & , iterasi = 1 \\ \sqrt{\frac{1}{nK}\sum_{i=1}^{n} w_i\, {e_i}^2} & , iterasi > 1 \end{cases}.$$

   dengan K = 0,199
   d. Menghitung nilai $u_i = \frac{e_i}{\hat{\sigma}_i}$
   e. Menghitung pembobot

$$w_i = \begin{cases} \left[1 - \left(\frac{u_i}{1{,}547}\right)^2\right]^2 & , |u_i| \le 1{,}547 \\ 0 & , |u_i| > 1{,}547 \end{cases}.$$

   f. Menghitung parameter $\hat{\beta}_S$ dengan metode WLS dengan pembobot w_i^0.
   g. Mengulangi langkah b-f sampai diperoleh nilai $\hat{\beta}_S$yang konvergen.
   h. Uji hipotesis untuk mengetahui apakah variabel bebas mempunyai pengaruh yang signifikan terhadap variabel tak bebas.

Estimasi-MM

Selain estimasi-M dan estimasi-S, estimasi *robust* yang lain adalah estimasi-MM. Estimasi-MM diperkenalkan oleh Yohai (1987).Metode ini berusaha untuk mempertahankan sifat *robust* dan resisten dari estimasi-S, serta sifat efisien dari estimasi-M.Prosedur estimasi ini adalah dengan mengestimasi parameter regresi menggunakan estimasi-S yang meminimumkan skala sisaan dari estimasi-M dan dilanjutkan dengan estimasi-M.Estimasi-MM bertujuan untuk

 ISBN : 978 – 979 – 16353 – 9 – 4

mendapatkan estimasi yang mempunyai nilai *breakdown* tinggi dan lebih efisien. Nilai *breakdown* adalah ukuran umum proporsi dari pencilan yang dapat ditangani sebelum pengamatan tersebut mempengaruhi model.Estimasi-MM merupakan penyelesaian dari

$$\hat{\beta}_{MM} = \sum_{i=1}^{n} x_{ij}\rho_1{}'\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta_j}{\widehat{\sigma}_s}\right) = 0 \quad (12)$$

dengan $y_i - \sum_{j=0}^{k} x_{ij}\beta_j$ adalah sisaan yang diperoleh dari estimasi parameter model regresi dengan estimasi-S dan $\widehat{\sigma}_s$ merupakan penyelesaian dari

$$\frac{1}{n}\sum_{i=1}^{n} \rho_0\left(\frac{y_i - \sum_{k=0}^{p} x_{ij}\widetilde{\beta_j}}{\widehat{\sigma}_s}\right) = K.$$

Algoritma penghitungan nilai estimasi-MM sebagai berikut:

1. Melakukan estimasi koefisien regresi pada data menggunakan MKT.
2. Menguji asumsi klasik dari model regresi.
3. Mendeteksi adanya pencilan dalam data.
4. Mengestimasi koefisien regresi*robust* menggunakan estimasi-MM.
   a. Menghitung nilai sisaan $e_i = y_i - \hat{y}_i$. dari estimasi-S
   b. Menghitung nilai $\widehat{\sigma}_i = \hat{\sigma}_{sn}$.
   c. Menghitung nilai $u_i = \frac{e_i}{\widehat{\sigma}_i}$.
   d. Menghitung pembobot
   e. $w_i = \begin{cases} \left[1 - \left(\frac{u_i}{4,685}\right)^2\right]^2 & , |u_i| \leq 4,685 \\ 0 & , |u_i| > 4,685 \end{cases}$.
   f. Menghitung parameter $\hat{\beta}_{MM}$ dengan metodeWLS dengan pembobot w_i^0.
   g. Mengulangi langkah b - e sampai diperoleh nilai $\hat{\beta}_{MM}$ yang konvergen.
   h. Uji hipotesis untuk mengetahui apakah variabel bebas mempunyai pengaruh yang signifikan terhadap variabel tak bebas.

C. OPTIMASI MODEL REGRESI

Pada penelitian ini data diambil dari BPS dan Kementerian Pertanian.Tidak semua faktor yang diduga mempengaruhi produksi kedelai tersedia untuk setiap propinsi. Data lengkap yang terkait dengan produksi kedelai untuk tahun 2011 adalahluas lahan panen, curah hujan, kelembaban, suhu, lama penyinaran, dan jumlah nilai tukar petani. Oleh karena itu akan dibahas model regresi linear produksi kedelai (Y) seluruh propinsi di Indonesia berdasarkanluas lahan panen (X_1), curah hujan (X_2), kelembaban (X_3), suhu (X_4), lama penyinaran (X_5), dan nilai tukar petani (X_6). Estimasi model regresi dengan MKT adalah

$$\hat{Y}_i = -14898 + 1{,}45\, X_1 + 13{,}2\, X_2 + 137X_3 - 481X_4 - 38\, X_5 + 137\, X_6 \quad (13)$$

dengan$R^2 = 99,7\ \%$, $R^2(adj)$= 99,6%, dan s= 4440,15

Kemudian dilakukan uji asumsi untuk melihat apakah asumsi regresi linear untuk model (13) dipenuhi atau tidak. Dari hasil uji asumsi, asumsi kenormalan, homoskedastik, dan nonautokorelas*i* ketiganya tidak dipenuhi, hanya asumsi nonmultikolinearitas yang dipenuhi, dan terdapat data pencilan yaitu data ke-11, 14, 15 dan 17.

Dilihat dari nilai F = 1188,77 dengan nilai-p = 0< 5%, ini menunjukkan bahwa model regresi linear Y dengan X_1 , X_2 ,X_3 , X_4 , X_5 dan X_6sudah baik. Selanjutnya karena ketiga asumsi tidak dipenuhi dan terdapat pencilan, maka akan dilakukan estimasi dengan model regresi linear *robust* yaitu dengan estimasi-M, estimasi-S, dan estimasi-MM.

Model Regresi dengan Estimasi-M

Model regresi linear produksi kedelai dengan estimasi-Madalah

$$\hat{Y}_i = 334481 + 1{,}46X_1 + 4{,}36X_2 + 30{,}6X_3 - 1292X_4 + 34{,}7X_5 - 48{,}4X_6 \quad (14)$$

dengan R^2 = 100 %, $R^2(adj)$= 100% dan s= 1016,53.

Model regresi (14) ini diperoleh dengan iterasi sebanyak 10 kalidan menunjukkan bahwa untuk peningkatan setiap satu hektar luas panen, satu mm curah hujan, satu persen kelembaban dan satu persen lama penyinaranmaka produksinya akan meningkat masing-masing sebesar 1,46; 4,36, 30,6; dan 34,7 ton. Untuk peningkatan, satu derajat Celsius suhu dan satu persen nilai tukar petani akan mengakibatkan penurunan produksi kedelai masing-masing sebesar 1.292 dan 48,4 ton.

Model (14) mempunyai data pencilan sebanyak dua yaitu data ke-7 dan 11. Selanjutnya dilakukan uji hipotesis untuk mengetahui apakah luas lahan panen, curah hujan, kelembaban, suhu, lama penyinaran, atau nilai tukar petani mempunyai pengaruh signifikan terhadap produksi kedelai di Indonesia tahun 2011.

Tabel 1 Hasil Uji Parsial Variabel BebasModel Prediksi Kedelai dengan Estimasi-M

	Koefisien	t	p	Kesimpulan
Konstan	33481	2,78	0,011	Signifikan
X_1	1, 46	248,71	0,000	Signifikan
X_2	4,36	1,25	0,225	Tidak signifikan
X_3	30,6	0,32	0,748	Tidak Signifikan
X_4	-1292	-4,41	0,000	Signifikan
X_5	34,7	1,23	0,230	Tidak signifikan
X_6	-48,4	-0,81	0,425	Tidak signifikan

Tabel 1 menghasilkan suatu kesimpulan yaitu luas lahan panen dan suhu memberikan pengaruh signifikanterhadap produksikedelai. Sedangkan curah hujan, kelembaban, lama penyinaran, dan nilai tukar petanitidak memberikan pengaruh yang signifikan terhadap produksi kedelai di Indonesia.

Model Regresi dengan Estimasi-S

Model regresi linear dengan estimasi-Sadalah

$$Y = -25768 + 1{,}43\,X_1 - 2{,}08\,X_2 + 177\,X_3 + 684\,X_4 + 1{,}8\,X_5 - 69{,}2\,X_6 \qquad (15)$$

dengan $R^2 = 99{,}9\%$, $R^2(adj) = 99{,}9\%$ dan $s = 287{,}345$.

Model regresi (15) menunjukkan bahwa untuk peningkatan setiap satu hektar luas panen, satu persen kelembaban, satu derajat Celsius suhu, dan satu persen lama penyinaran, maka produksinya juga meningkat masing-masing sebesar 1,43; 177; 684; dan 1,8 ton. Untuk peningkatan satu mm curah hujan dan satu persen nilai tukar petani akan mengakibatkan penurunan produksi kedelai masing-masing sebesar 2,08 dan 69,2 ton.

Model (15) tidak mempunyai pencilan. Selanjutnya dilakukan uji hipotesis untuk mengetahui apakah luas lahan panen, curah hujan, kelembaban, suhu, lama penyinaran, atau nilai tukar petani mempunyai pengaruh signifikan terhadap produksi padi di Indonesia tahun 2011.

Tabel 2 Hasil Uji Parsial Variabel Bebas Model Prediksi Kedelai dengan Estimasi-S

	Koefisien	t	p	Kesimpulan
Konstan	-25768	-2,78	0,017	Signifikan
X_1	1,43	134,29	0,000	Signifikan
X_2	-2,08	-1,47	0,164	Tidak signifikan
X_3	177	4,98	0,000	Signifikan
X_4	684	2,38	0,032	Signifikan
X_5	1,8	0,17	0,871	Tidak signifikan
X_6	-69,2	-3,69	0,002	Signifikan

Tabel 2 menghasilkan suatu kesimpulan bahwa empat variabel bebas yaitu luas lahan panen, kelembaban, suhu dan nilai tukar petani mempunyai pengaruh yang signifikan, sedangkan

 ISBN : 978 – 979 – 16353 – 9 – 4

curah hujan dan lama penyinaran tidak memberikan pengaruh signifikan terhadap produksi kedelai di Indonesia.

Model Regresi dengan Estimasi-MM

Model regresi linear dengan estimasi-MMadalah

$$\hat{Y}_i = -\ 27128 + 1{,}43X_1 + 0{,}62X_2 + 119\ X_3 + 768\ X_4 - 0{,}7\ X_5 - 37{,}4\ X_6 \qquad (16)$$

dengan R^2= 99,6%, $R^2(adj)$= 99,5% dan s= 697,055.

Model regresi (16) menunjukkan bahwa untuk peningkatan setiap satu hektar luas panen,satu mm curah hujan, satu persen kelembaban dan satu derajat Celsius suhu maka produksinya juga meningkat masing-masing sebesar 1,43; 0,62; 119 dan 768 ton. Untuk peningkatan satu persen lama penyinaran dan satu persen nilai tukar petani akan mengakibatkan penurunan produksi kedelai masing-masing sebesar 0,7 dan 37,4 ton

Model (16) mempunyai satu data pencilan yaitu data ke-7. Setelah diperoleh model regresi linear (16), dilakukan uji hipotesis untuk mengetahui apakah luas lahan panen, curah hujan, kelembaban, suhu, lama penyinaran, atau nilai tukar petani mempunyai pengaruh signifikan terhadap produksi kedelai di Indonesia tahun 2011.

Tabel 3 Hasil Uji Parsial Variabel BebasModel Prediksi Kedelai dengan Estimasi-MM

	Koefisien	t	p	Kesimpulan
Konstan	-27128	-1,59	0,130	tidak signifikan
X_1	1,43	57,43	0.000	signifikan
X_2	0,62	0,22	0,828	tidak signifikan
X_3	119	1,78	0,092	tidak signifikan
X_4	768	1,38	0,183	tidak signifikan
X_5	-0,7	-0,03	0,974	tidak signifikan
X_6	-37,4	-0,90	0,380	tidak signifikan

Tabel 3 menghasilkan suatu kesimpulan bahwa hanya satu variabel bebas luas lahan panen yang memberikan pengaruh yang signifikan terhadap produksi kedelai, sedangkan lima variabel bebas lainnya yaitu curah hujan, kelembaban, suhu, lama penyinaran, dan nilai tukar petani tidak memberikan pengaruh yang signifikan terhadap produksi kedelai di Indonesia.

Model Optimal

Ada beberapa kriteria yang dapat digunakan untuk menentukan model regresi terbaik, yaitu koefisien determinasi R^2atau $R^2_{adjusted}$dan deviasi standar *s.* Model terbaik akan mempunyai R^2 atau $R^2_{adjusted}$ terbesar dan *s* terkecil.

Tabel 4 Nilai $R,^{2,}$ R^2 *adjusted* , s, Variabel yang Signifikan danPencilan untuk model (14) - (16)

	Estimasi-M	Estimasi-S	Estimasi-MM
Model	(14)	(15)	(16)
R^2	99,7 %	99,9 %	99,6 %
R^2 *adjusted*	99,6 %	99,9 %	99,5 %
s	1016,53	287,345	697,055
Variabel yang signifikan	X_1, X_4	X_1, X_3, X_4, X_6	X_1
Data pencilan, data ke-	7, 11	-	7

Dari Tabel 4 terlihat bahwa model dengan estimasi-S memberikan R^2, R^2*adjusted* yang lebih besar dan *s* yang lebih kecil daripada model dengan estimasi-M atau MM, sehingga model terbaiknya adalah model dengan estimasi-S, yaitu model (15).Selain itu pada model ini tidak terdapat pencilan.

Dalam model ini empat variabel bebas yaitu luas lahan panen, kelembaban, suhu dan nilai tukar petani memberikan pengaruh signifikanterhadap produksi kedelai di Indonesia. Sedangkan curah hujan dan lama penyinaran tidak memberikan pengaruh yang signifikan terhadap produksi kedelai di Indonesia.Selanjutnya akan ditentukan model kedelai dengan estimasi-S dengan variabel bebas yang signifikan (luas lahan panen (X_1), kelembaban (X_3), suhu (X_4) dan nilai tukar petani (X_6)), hasilnya adalah

1. Model kedelai untuk variabel yang signifikan dengan MKT adalah
$$Y = -28256+ 1{,}45\, X_1 + 309\, X_3 - 505\, X_4 + 141 X_6 \quad (17)$$
dengan R^2 = 99,6 % , $R^2(adj)$= 99,6 % dan s= 4382,5
2. Model kedelai untuk variabel yang signifikan dengan estimasi-S adalah
$$Y = -50896+ 1{,}43\, X_1+ 190\, X_3 + 878\, X_4 + 106\, X_6 \quad (18)$$
dengan R^2 = 99,9 % , $R^2(adj)$= 99,9 % dan s= 341,375.

Untuk model (18), $R^2(adj)$= 99,9 % menunjukkan bahwa variasi total Y sebesar 99,9% diterangkan oleh X_1, X_3, X_4, dan X_6, sedangkan yang 0,1 % disebabkan oleh variabel yang lain. Selanjutnya dilakukan uji signifikansi model, diperoleh nilai F = 4695,70 dengan nilai-p = 0< 5 %, sehingga dapat disimpulkan bahwa model sudah baik. Uji parsial pada Tabel 5 menunjukkan bahwa semua variabel bebas pada model (18) signifikan.

Tabel 5. Hasil Uji Parsial Variabel Bebas Model Prediksi Kedelai dengan Estimasi-S

	Koefisien	t	p	Kesimpulan
Konstan	- 50896	-5,5	0.000	Signifikan
X_1	1,43	129,17	0.000	Signifikan
X_3	190	5,95	0.000	Signifikan
X_4	878	2,85	0.011	Signifikan
X_6	106	2,81	0.012	Signifikan

D. KESIMPULAN

Model regresi optimal yang diperoleh untukprediksi produksi kedelai di Indonesia adalah model dengan estimasi-S yaitu$Y$ = -50896+ 1,43 X_1 + 190 X_3 + 878 X_4 + 106 X_6.Variasi total Y sebesar 99,9 % diterangkan oleh X_1, X_3, X_4, dan X_6, sedangkan yang 0,1 % disebabkan oleh variabel yang lain. Setiap peningkatan satu hektar luas lahan panen, satu persen kelembaban, satu derajat Celcius suhu, dan satu persen nilai tukar petani masing-masing akan meningkatkan produksi kedelai berturut-turut sebesar 1,43; 190; 878; dan 106 ton.

E. DAFTAR PUSTAKA

Badan Pusat Statistik, 2012,*Production of Paddy, Maizze and Soybeans*, www.bps.go.id/releases/Production of Paddy Maizze and Soybeans.

Chen, C., 2002, Robust regression and Outlier Detection with the ROBUSTREG Procedure,Paper 265-27, *Statistics and Data Analysis*, SUGI 27, North Carolina:SAS Institute Inc.

Draper,N.R and Smith, 1998,*Applied Regression Analysis*, Third Edition, United States: Wiley Intercience Publication.

Montgomery, D.C. and Peck, E.A., 2006, *Introduction to Linear Regression Analysis*, New York:John Wiley & Sons Inc.

Rousseeuw, P.J. and Yohai, V.J., 1984.Robust Regression by Mean of S-Estimators, *Robust and Nonlinear Time Series*, eds. J. Franke, W. Hardle, and D. Martin, *Lecture Notes in Statistics*, 26, 256 – 272, *Berlin:* Springer-Verlag.

Susanti, Y. dan Pratiwi, H., 2011, Robust Regression Model for Predicting the Soybean Production in Indonesia, *Canadian Journal on Scientific and Industrial Research*, Vol. 2 No. 9, December 2011, 318-328.

Susanti, Y. dan Pratiwi, H., 2012. Modelling of Soybean Production in Indonesia Using Robust Regression, *Bionatura, Jurnal Ilmu-Ilmu Hayati dan Fisik*, Vol. 14 No. 2, Juli 2012, 148-155.

Susanti, Y., Pratiwi, H. dan Liana, T., 2009, Application of M-estimation to Predict Paddy Production in Indonesia, dipresentasikan pada *IndoMS International Conference onMathematics and Its Applications (IICMA)*, Yogyakarta.

Yohai, V.J., 1987. High Breakdown Point and High Efficiency Robust Estimates for Regression, *The Annals of Statistics*, Vol. 15, 642-656.

Yuliana dan Susanti, Y., 2008, Estimasi-M dan sifat-sifatnya pada Regresi Linear Robust, *Jurnal Math-Info*, Vol. l, Surakarta: UNS.

Yuliana dan Susanti, Y.,2008, Estimasi-M dan Sifat-sifatnya pada Regresi Linear Robust, *Jurnal Math-Info*, Vol. l No. 10, Surakarta: UNS.