PROSIDING SEMINAR NASIONAL STATISTIKA
UNIVERSITAS DIPONEGORO 2013
ISBN: 978-602-14387-0-1

APLICATION OF M-ESTIMATION FOR RESPONSE SURFACE MODEL WITH DATA OUTLIERS

Edy Widodo[1], Suryo Guritno[2] and Sri Haryatmi[2]
[1]Doctoral Student of Mathematics UGM,
[2]Department of Mathematics UGM
edykafifa@gmail.com

Abstract

Relationship between the response variable and the independent variables in a limited area of operation, in the Response Surface Methodology, used a second order polynomial function. The parameters of the model are usually estimated by Least Squares Method. However, this method is very sensitive to outliers. Outliers can affect the results of statistical analysis, as outliers are very likely to produce a large residual and often affect the regression models generated. Thus, the resulting model estimates to be biased and result in errors in the determination of the actual optimal point. Therefore, it takes a strong response surface model/robust against outliers. Proposed as an alternative to using the M-Estimation, for estimating the parameters in the response surface model. In this paper will be shown in the application of M-Esimation Response Surface Model.

Keywords: Response Surface Models, Estimation M, Outliers

1. Pendahuluan

Metode Permukaan Respon (MPR*)* adalah suatu kumpulan dari teknik–teknik statistika dan matematika atau metode yang berguna untuk menganalisis permasalahan tentang beberapa variabel bebas yang mempengaruhi variabel tak bebas atau respon, dan bertujuan mengoptimalkan respon (maksimum, minimum atau lebih luas lagi, mencari kondisi disekitar titik stasioner yang mengandung *ridge*). MPR dapat digunakan untuk mencari suatu fungsi pendekatan yang cocok untuk meramalkan respon yang akan datang dan menentukan nilai–nilai dari prediktor yang mengoptimalkan respon. MPR pertama kali diperkenalkan oleh *Box dan Wilson (1951),*

Secara umum ada tiga tahapan utama dalam MPR, yaitu: 1) **pengumpulan data**, melalui pemilihan strategi rancangan percobaan yang tepat, 2) **estimasi model/pemodelan data**, melalui pemilihan metode pemodelan regresi yang tepat, dan 3) **optimasi**, melalui pemilihan metode optimasi yang akan digunakan untuk mengidentifikasi pengaturan dari peubah bebas yang mengoptimalkan peubah respon.

Ketiga tahapan tersebut saling terkait, dimana setiap tahapan akan mempengaruhi tahapan berikutnya. Oleh karena luasnya bidang kerja dari MPR, maka

ISBN: 978-602-14387-0-1

penelitian ini akan difokuskan pada estimasi model regresi untuk MPR, dengan asumsi bahwa rancangan percobaan sudah memuaskan dan data telah dikumpulkan

Untuk mendekati hubungan antara peubah respon dan peubah bebas pada daerah asal (daerah operasi) yang terbatas, dalam MPR, digunakan fungsi polinomial orde dua (model kuadratik). Model ini selanjutnya disebut dengan model orde dua. Dalam bentuk matrik, model orde dua dapat dituliskan sebagai berikut :

$$\boldsymbol{E(y) = \beta_0 + x'\beta + x'Bx} \tag{1}$$

dengan :

$\boldsymbol{x = (x_1, x_2, \ldots, x_k)}$

$\beta = (\beta_1, \beta_1, \ldots, \beta_k)'$, dan

$$\boldsymbol{B} = \begin{pmatrix} \beta_{11} & \beta_{12}/2 & \ldots & \beta_{1k}/2 \\ \beta_{21}/2 & \beta_{22} & \ldots & \beta_{2k}/2 \\ . & . & \ldots & . \\ \beta_{k1}/2 & \beta_{k2}/2 & \ldots & \beta_{kk} \end{pmatrix} \tag{2}$$

Selanjutnya untuk melakukan estimasi terhadap koefisien-koefisien regresi linier sebagaimana dalam Persamaan (1), dalam model permukaan respon biasanya digunakan Metode Kuadrat Terkecil (MKT). Namun, metode ini sangat peka terhadap adanya penyimpangan asumsi pada data. Salah satu asumsi penting dalam analisis regresi adalah asumsi sebaran normal (normalitas). Asumsi normalitas seringkali dilanggar saat data mengandung *outliers.*

Menurut *Xu., J. (2006), outliers* dapat mempengaruhi estimasi parameter dalam model. *Outliers* dapat berpengaruh terhadap hasil analisis statistik, karena *outliers* sangat mungkin menghasilkan residual yang besar dan sering berpengaruh terhadap model regresi yang dihasilkan. Sehingga, jika model ini diterapkan akan menjadi masalah, karena akan menyebabkan estimasi model yang dihasilkan menjadi bias dan akan berakibat pada kesalahan dalam penentuan titik optimal yang sebenarnya. Oleh karena itu, dibutuhkan suatu model permukaan respon yang kuat/kekar terhadap *outliers*.

Beberapa macam cara yang digunakan oleh peneliti untuk memperlakukan atau menindaklanjuti adanya *outliers* pada data. Yang pertama **mengeluarkan *outliers***, dengan pertimbangan bahwa ada kemungkinan *outliers* tersebut disebabkan oleh kekeliruan, sehingga bukan data yang sebenarnya. Selain itu pengeluaran *outliers* tidak

ISBN: 978-602-14387-0-1

mengurangi informasi dan dianggap sudah bisa terwakili oleh sebagian besar data lainnya. Yang kedua **menggunakan data secara keseluruhan atau tanpa mengeluarkan *outliers***, tetapi dengan memberikan bobot yang kecil untuk data *outliers*, metode ini selanjutnya dikenal dengan nama metode regresi *robust.*

Regresi *robust* merupakan metode regresi yang digunakan ketika ada beberapa *outliers* pada model. Metode ini merupakan alat penting untuk menganalisa data yang dipengaruhi oleh *outliers* sehingga dihasilkan model yang *robust* atau kekar terhadap *outliers*. Suatu estimator yang kekar adalah relatif tidak terpengaruh oleh perubahan besar pada bagian kecil data atau perubahan kecil pada bagian besar data. Salah satu metode regresi *robust* diperkenalkan adalah penaksir-M *(Maximum Likelihood type).*

Penaksir-M adalah metode yang paling sederhana dan paling banyak digunakan serta mempunyai nilai efisiensi yang tinggi. Wen Wan (2007), merekomendasikan untuk mengembangkan metode yang tahan terhadap *outliers*, dengan merujuk pada hasil penelitian Assaid (1997). Di dalam penelitianya, *Assaid (1997),* telah mengusulkan suatu metode yang dinamakan dengan *Outlier Resistant Model Robust Regression (ORMRR).* Usulan *ORMRR* dilatarbelakangi oleh keinginan untuk mendapatkan sebuah rumusan taktis dalam menangani *misspecification model* dengan adanya *outliers*, pada kasus-kasus di luar model permukaan respon.

Dengan memperhatikan latar belakang di atas dalam makalah ini akan ditunjukkan tentang penggunaan penaksir-M untuk model permukaan respon dengan data *outliers.*

2. Penaksir –M

Penaksir–M merupakan metode regresi *robust* yang sering digunakan karena dipandang baik untuk mengestimasi parameter yang disebabkan oleh *outliers*. Metode robust ini menduga koefisien regresi dari data yang mengandung *outliers* dengan meminimumkan sebuah fungsi objektif (ρ) dari galat

$$\min_{\hat{\beta}} \sum_{i=1}^{n} \rho(e_i) = \min_{\hat{\beta}} \sum_{i=1}^{n} \rho\left(y_i - x_i^T \hat{\beta}\right) \qquad (3)$$

ISBN: 978-602-14387-0-1

Penaksir-M tidak perlu *scale-invariant* dalam artian jika galat dikuadratkan dengan sebuah konstanta, maka persamaan (3) berubah. Untuk memperoleh *scale-invariant* untuk jenis estimator ini, solusinya adalah:

$$\min_{\hat{\beta}} \sum_{i=1}^{n} \rho\left(\frac{e_i}{s}\right) = \min_{\hat{\beta}} \sum_{i=1}^{n} \rho\left(\frac{y_i - x_i^T\hat{\beta}}{s}\right) \tag{4}$$

dengan $s = \frac{median|e_i - median(e_i)|}{0.6745}$ dan $\rho\left(\frac{e_i}{s}\right)$ adalah fungsi simetris dari residual atau fungsi yang memberikan kontribusi pada masing-masing residual pada fungsi objektif.

Selanjutnya jika $\varphi = \rho'$ adalah derivative dari ρ, maka untuk meminimumkan persamaan (4) adalah:

$$\sum_{i=1}^{n} x_{ij}\varphi\left(\frac{y_i - x_i^T\hat{\beta}}{s}\right) = 0\,; j = 1,2,3,\dots,m \tag{5}$$

$\varphi(.)$ merupakan fungsi *influence* yang umumnya berbentuk non-linier, sehingga persamaan (5) dapat diselesaikan dengan metode iterasi yaitu dengan metode *iterasi reweighted least square(IRLS)* yang menghasilkan persamaan:

$$\sum_{i=1}^{n} x_{ij}\varphi\left(\frac{y_i - x_i^T\hat{\beta}}{s}\right) = \sum_{i=1}^{n} x_{ij}\frac{\varphi\left(\frac{y_i - x_i^T\hat{\beta}}{s}\right)}{\left(\frac{y_i - x_i^T\hat{\beta}}{s}\right)}\left(\frac{y_i - x_i^T\hat{\beta}}{s}\right) = 0\,, j = 0,1,\dots,q \tag{6}$$

atau

$$\sum_{i=1}^{n} x_{ij}w_{io}\left(\frac{y_i - x_i^T\hat{\beta}}{s}\right)\,; j = 0,1,..,q \tag{7}$$

dengan

$$w_{io} = \begin{cases} \dfrac{\varphi\left(\frac{y_i - x_i^T\hat{\beta}}{s}\right)}{\left(\frac{y_i - x_i^T\hat{\beta}}{s}\right)} & jika\ y_i \neq x_i^T\hat{\beta} \\ 1 & ;jika\ y_i = x_i^T\hat{\beta} \end{cases} \tag{8}$$

dan dalam notasi matriks menjadi:

$$X^T W_0 X\hat{\beta} = X^T W_0 y \tag{9}$$

ISBN: 978-602-14387-0-1

dengan W_0 merupakan diagonal matriks *"weight"* yang berukuran n x n dengan diagonal elemen $w_{10}, w_{20}, \ldots, w_{n0}$ yang diberikan oleh persamaan (8). Regresi terboboti tersebut dapat digunakan sebagai alat untuk mendapatkan penaksir-M. Sehingga estimasi parameter menjadi:

$$\hat{\beta} = (X^T W_0 X)^{-1} X^T W_0 y \qquad (10)$$

Prosedur penaksiran dengan menggunakan penaksir-M diuraikan sebagai berikut :

a) dihitung penaksir **β,** dinotasikan **b** menggunakan MKT, sehingga didapatkan $\hat{y}_{i,0}$ dan $\varepsilon_{i,0} = y_i - \hat{y}_{i,0}$, ($i$ = 1, 2, ... n) yang diperlakukan sebagai nilai awal, dimana y_i adalah hasil eksperimen.

b) dari nilai-nilai residual ini dihitung $\hat{\sigma}_0$, dan pembobot awal $w_{i,0} = \dfrac{\psi(\varepsilon^*_{i,0})}{(\varepsilon^*_{i,0})}$. Nilai $\psi(\varepsilon_i^*)$ di hitung sesuai fungsi Huber, dan $\varepsilon_{i,0}^* = \varepsilon_{i,0} / \hat{\sigma}_0$.

c) Disusun matrik pembobot berupa matrik diagonal dengan elemen $w_{1,0}$, $w_{2,0}$, . . . , $w_{n,0}$, dinamai W_0.

d) Dihitung penaksir koefisien regresi, $\mathbf{b}_{\text{Robust ke 1}} = (X^T W_0 X)^{-1} X^T W_0 Y$

e) Dengan menggunakan $\mathbf{b}_{\text{Robust ke 1}}$ dihitung pula $\sum_{i=1}^{n} | y_i - \hat{y}_{i,1} |$ atau $\sum_{i=1}^{n} | \varepsilon_{i.1} |$

f) Selanjutnya langkah b) sampai dengan e) diulang sampai didapatkan $\sum_{i=1}^{n} | \varepsilon_{i.m} |$ konvergen.

3. Ketepatan Model

Dalam MPR terdapat dua kesalahan yaitu kesenjangan terhadap ketepatan model (*lack of fit/lof*) dan kesalahan murni (*pure error/pe*). *Lof* adalah suatu ukuran untuk mengetahui ketepatan murni atau dengan kata lain jika diketahui bahwa model harusnya linier namun dipaksakan lengkung maka terjadi kesenjangan. Dalam *lof* ada syarat yang harus dipenuhi yaitu dalam percobaan itu harus ada perulangan.

4. Pencilan *(Outliers)*

Johnson (2007) mendefinisikan *outliers* sebagai suatu pengamatan yang tak konsisten dengan kumpulan pengamatan lainnya. Selain itu *outliers* dapat didefinisikan sebagai suatu pengamatan yang menyimpang dari kumpulan pengamatan

ISBN: 978-602-14387-0-1

lainnya, sehingga tidak mengikuti sebagian besar pola. Akibatnya letaknya jauh dari pusat pengamatan

Beberapa definisi outliers yang lain diuraikan oleh Soemartini (2007), yang mengambil dari beberapa pendapat diantaranya adalah 1) Ferguson yang mendefinisikan pencilan sebagai suatu pengamatan yang menyimpang dari sekumpulan pengamatan yang lain; 2) Barnett yang mendefinisikan *outliers* sebagai pengamatan yang tidak mengikuti sebagian besar pola dan terletak jauh dari pusat pengamatan; dan 3) Sembiring yang mendefinisikan *outliers* sebagai pengamatan yang jauh dari pusat data yang mungkin berpengaruh besar terhadap koefisien regresi.

Outliers secara umum merupakan satu atau sebagian kecil data yang letaknya jauh dari pola kumpulan data secara keseluruhan. *Outliers* dalam sekumpulan data hasil pengamatan adalah sebuah pengamatan yang muncul dan nilainya tidak konsisten dengan nilai data pengamatan yang lainnya. Pengaruh *outliers* dalam analisis data dapat dibedakan berdasarkan asal *outliers* tersebut, yaitu yang berasal dari peubah respon (y-*outliers;* titik *influence*) atau berasal dari peubah bebasnya (*x-outliers; titik leverage*).

Sebuah pengamatan yang berbeda dari sekumpulan data lainnya, dapat berpengaruh besar pada analisis regresi. *Outliers* dapat menyebabkan hal–hal berikut : 1) residual yang besar dari model yang terbentuk atau $E[e] \neq 0$, 2) varians pada data tersebut menjadi lebih besar dan 3) taksiran interval akan memiliki rentang yang lebar

Terdapat beberapa jenis *outliers* dalam suatu data, yaitu 1) *Good leverage* yaitu pengamatan yang berada di ruang distribusi tetapi sudah tidak berada di daerah mayoritas data, 2) *Bad leverage* yaitu pengamatan yang tidak berada baik dalam ruang distribusi pengamatan maupun daerah mayoritas data dan 3) *Pencilan orthogonal* yaitu pengamatan yang mempunyai jarak yang sangat besar dari daerah mayoritas data sehingga pengamatan tersebut sudah tidak dapat dilihat dalam ruang distribusinya.

5. Metode Penelitian

Dalam makalah ini akan ditunjukkan dengan simulasi komputer dengan menggunakan paket software Minitab 16. Data yang digunakan adalah data *xanthan gum production* yang dilakukan oleh Psomas dkk. (2007). Data yang digunakan meliputi *xanthan gum production* (y) sebagai variabel respon, dengan variabel

ISBN: 978-602-14387-0-1

independen meliputi a*gitation rate* (x_1), *temperature* (x_2), *and time of cultivation* (x_3) yang diperoleh dengan menggunakan percobaan *central composit design.* Data selengkapnya dapat dilihat pada tabel berikut ini

Tabel 1. Data *Xanthan Gum Production*, (Psomas dkk. (2007))

Run	Agitation rate (x1)	Temperature (x2)	Time (x3)	x_1	x_2	x_3	y
1	100	25	24	-1	-1	-1	0.278
2	600	25	24	1	-1	-1	0.375
3	100	35	24	-1	1	-1	0.141
4	600	35	24	1	1	-1	0.333
5	100	25	72	-1	-1	1	0.315
6	600	25	72	1	-1	1	0.692
7	100	35	72	-1	1	1	0.279
8	600	35	72	1	1	1	0.699
9	100	30	48	-1	0	0	0.215
10	600	30	48	1	0	0	0.486
11	350	25	48	0	-1	0	0.583
12	350	35	48	0	1	0	0.569
13	350	30	24	0	0	-1	0.348
14	350	30	72	0	0	1	0.511
15	350	30	48	0	0	0	0.503
16	350	30	48	0	0	0	0.467
17	350	30	48	0	0	0	0.453
18	350	30	48	0	0	0	0.475 **(4.75) (0.0475)**

Keterangan : Tanda kurung pada data y adalah simulasi untuk data *outliers* (pencilan)

Tahapan penelitian

a) dilakukan estimasi model orde dua dengan menggunakan metode kuadrat terkecil pada data asli
b) dilakukan estimasi model orde dua dengan menggunakan penaksir-M pada data asli
c) dilakukan estimasi model orde dua dengan menggunakan metode kuadrat terkecil pada data simulasi (yang mengandung outlier, yaitu dengan mengubah 0.475 menjadi 4.75 dan 0.0475)
d) dilakukan estimasi model orde dua dengan menggunakan penaksir-M pada data simulasi (yang mengandung outlier, yaitu dengan mengubah 0.475 menjadi 4.75 dan 0.0475)
e) membandingkan hasil estimasi model dari tahap a) sampai dengan d), untuk menentukan model terbaik

ISBN: 978-602-14387-0-1

6. Hasil dan Pembahasan

Simulasi dilakukan dengan mengambil salah satu titik pusat untuk diberikan perlakuan, yaitu dengan mengubah nilai asli pada data tersebut dengan nilai tertentu. Sedangakan titik-titik lain tidak diubah. Dalam simulasi ini titik yang terpilih adalah 0.475. Dari titik yang terpilih ini selanjutnya dilakukan simulasi dengan tiga kondisi yaitu 0.475, 4.75 dan 0.0475. Ringkasan hasil simulasi yang telah dilakukan dapat dilihat pada Tabel 2.

Tabel 2. Hasil Simulasi Data

	Tanpa outlier (0.475)		**Ada outlier (4.75)**		**Ada outlier (0.0475)**	
	MKT	**Penaksir M**	**MKT**	**Penaksir M**	**MKT**	**Penaksir M**
MSE	0.000429	**0.000382**	1.928	**0.4735**	0.02005	**0.003804**
MSE_{LOF}	0.000420	**0.000399**	0.343	**0.0219**	0.00449	**0.000710**
MSE_{pe}	0.000444	**0.000354**	4.571	**1.2261**	0.04599	**0.008960**
Titik stasioner	-0.34216 0.96605 -4.10476	-0.34250 0.95335 -4.08870	-0.02147 -0.29610 -2.02700	1.4609 -9.1872 14.2820	-0.31633 0.65200 -3.47154	-0.33539 0.87460 -3.93138
Titik optimum	0.234442	0.233917	1.03789	1.57088	0.205845	0.229084

Keterangan : MSE = Mean Square Error; LOF : lack of fit; pe : pure error

Dari Tabel 2, pada kasus tanpa ada outlier terlihat bahwa penaksir MKT dan penaksir-M mempunyai nilai MSE, MSE_{LOF} , dan MSE_{pe} yang hampir sama bahkan ada kecenderungan bahwa nilai-nilai pada pada penaksir-M mempunyai nilai yang lebih kecil. Hal ini menunjukkan bahwa kedua metode penaksir mempunya kemampuan yang relatif sama pada kasus data hasil percobaan yang tidak memuat outlier. Pada kasus di atas penaksir-M mempunyai kemampuan yang lebih unggul karena memiliki nilai MSE yang lebih rendah dibandingkan dengan penaksir MKT.

Pada kasus data hasil percobaan yang memuat *outliers* penaksir-M memberikan hasil yang lebih baik dibandingkan dengan MKT hal ini terlihat dari nilai MSE untuk penaksir-M yang memberikan hasil yang lebih kecil dibandingkan dengan penaksir MKT. Untuk kasus dimana terdapat outlier 4.75 yaitu dengan mengubah data asli dari 0.475 menjadi 4.75 pada hasil pengamatan, MSE, MSE_{LOF} , dan MSE_{pe} penaksir-M memberikan nilai yang lebih kecil dibandingkan dengan penaksir MKT. Hal yang sama

ISBN: 978-602-14387-0-1

juga terjadi pada kasus data pengamatan yang diubah dari 0.475 menjadi 0.0475, penaksir-M juga masih memberikan nilai yang lebih kecil dibandingkan dengan penaksir MKT. Hal ini menunjukkan bahwa penaksir-M mempunyai tingkat kesalahan yang lebih kecil dibandingkan dengan penaksir MKT, terutama untuk data yang memuat *outlier*s.

7. Kesimpulan

Telah ditunjukkan bahwa untuk kasus data hasil percobaan pada metode permukaan respon yang mengandung *outliers*, penaksir M dapat diterapkan untuk menaksir model permukaan respon.

DAFTAR PUSTAKA

Assaid, C., 1997, Outlier Resistant Model Robust Regression, *Ph.D. Dissertation*, Department of Statistics, Virginia Polytechnic Institute & State University, Blacks-burg, VA

Box, G.E.P. dan Wilson, K.B., 1951, On the experimental attainment of optimum conditions, *Journal of the Royal Statistical Society Series B*, 13, 1-45.

Huber, P.J.,1981, *Robust Statistics*, Wiley, New York.

Johnson, R.A., dan Wichern, D.W., 2007, *Applied multivariate statistical analysis*, sixth edition, Prentice Hall, New Jersey

Psomas, S.K., Liakopoulou-Kyriakides, M., and Kyriakidis, D. A. (2007). Optimization Study of Xanthan Gum Production Using Response Surface Methodology. *Biochemical Engineering Journal*, *Vol.35, pp. 273-280.*

Soemartini. (2007). "Outlier (Pencilan)". Bandung: UNPAD

Xu, J., Abraham, B., dan Steiner, S.H., 2006, Outlier Detection Methods in Multivariate Regression Models, *http://www.bisrg.uwaterloo.ca/archive/RR-06-07.pdf*, diakses 15 oktober 2012.

Wan W., 2007. Semi-parametric techniques for multi-response optimization, *PhD Dissertation*, Department of Statistics, Virginia Polytechnic Institute & State University, Blacksburg, VA

Widodo, E., Guritno, S., dan Haryatmi, S. 2013. Penaksir M untuk model permukaan respon dengan data outliers, seminar nasional Statistika UII Yoyakarta