

METODE REGRESI ROBUST PENDUGA WELSCH UNTUK MENGATASI MASALAH PENCILAN

Dian Cahyawati S.

Abstrak : Metode Kuadrat Terkecil (MKT) digunakan sebagai penduga parameter pada analisis regresi. Metode ini memerlukan beberapa asumsi klasik yang harus dipenuhi yaitu kenormalan, kehomogenan ragam dan kebebasan bagian sisaan pada model yang terbentuk. Pengujian asumsi-asumsi itu dilakukan dengan menggunakan metode diagnosis sisaan model. Beberapa metode diagnosis sisaan yang dapat digunakan adalah nilai t -residual dan atau DFITS. Nilai t -residual atau juga DFITS dari suatu titik data yang relatif besar, menunjukkan bahwa titik data itu dapat dicurigai sebagai pencilan. Adanya data pencilan (outliers) merupakan salah satu penyebab asumsi klasik tidak terpenuhi. Sehingga untuk melakukan pendugaan parameter pada data yang mengandung pencilan, perlu digunakan metode lain yaitu metode regresi robust. Salah satunya adalah penduga Welsch. Hasil penelitian terhadap tiga macam data yang terdiri atas satu, dua, dan tiga variabel bebas, yang mengandung pencilan, menunjukkan bahwa penggunaan regresi robust penduga Welsch memberikan nilai-nilai penduga parameter yang lebih baik dari MKT. Hal ini berdasarkan nilai standar error hasil penduga Welsch memberikan nilai yang lebih kecil. Selain diperoleh penduga parameter yang lebih baik, penerapan regresi robust penduga Welsch juga tidak menghilangkan informasi apapun dari keseluruhan data amatan seperti pada MKT saat pencilannya dibuang terlebih dahulu.

Kata Kunci : Metode kuadrat terkecil, nilai t - residual penduga Welsh

Abstract : The usage of Ordinary Least Square (OLS) method as parameter estimator on regression analysis, oblige fulfilled of some classic assumptions that is normality, homogeneous of variance and independency term of residual at formed regression model. If that assumptions do not fulfilled, hence the usage of OLS method will yield bias parameter estimate, and inappropriate to be used. One of the cause do not fulfill of that classic assumptions is the existence of data outliers at observation data. Hence, to conduct parameter estimate on data which have outliers require other method to be used that is robust regression method. One of the parameter estimator in robust regression is Welsch estimation.

The result of the research to three kinds of data which have outliers data, that is data with one, two, and three independent variables, indicating that parameter estimate values by Welsch estimation in robust regression better than parameter estimate values by OLS method. This is based on the estimate of standard error its parameter estimation value. Besides of better parameter estimation values, applying of Welsch estimation in robust regression do not so eliminate any information from whole observation data like at OLS method with eliminating outliers data beforehand.

Key word: OLS, Independency term of residual, Welsch estimation.

PENDAHULUAN

Metode Kuadrat Terkecil (MKT) digunakan untuk mendapatkan nilai-nilai penduga parameter model regresi. Penggunaan MKT memerlukan beberapa asumsi klasik yang harus dipenuhi, yaitu kenormalan, kehomogenan ragam, dan kebebasan komponen sisaan. Dengan asumsi-asumsi klasik itu, maka penduga parameter yang diperoleh akan bersifat *Best Linear Unbiased Estimator (BLUE)*.

Tetapi dalam berbagai kasus tidak jarang ditemui hal-hal yang menyebabkan tidak terpenuhinya asumsi-asumsi itu. Saat asumsi ada yang tidak terpenuhi, maka penggunaan MKT akan memberikan kesimpulan yang bersifat kurang baik atau berbias, interpretasi hasil yang diberikan juga menjadi tidak valid.

Salah satu penyebab tidak terpenuhinya asumsi klasik adalah adanya pencilan (*outliers*) pada data amatan. Pencilan ini dapat diketahui dengan melakukan diagnosis sisaan terhadap model regresi yang terbentuk. Untuk mengatasi pelanggaran asumsi karena data pencilan, diperlukan metode analisis lain yang dapat mengatasi adanya pencilan dalam data amatan, yaitu metode regresi robust, salah satunya adalah penduga Welsch.

Diharapkan, melalui metode regresi robust, dapat dihasilkan kesimpulan yang lebih baik dari MKT, atau akan diperoleh

hasil yang sama baik dan validnya dengan MKT saat asumsi klasik dipenuhi.

TINJAUAN PUSTAKA

Dalam proses pembentukan model, pola umum data amatan dapat diketahui dan dilihat dari plot sebaran titik-titik data. Sedangkan dalam proses analisis data, sangat diperlukan untuk melihat ada tidaknya penyimpangan dari pola umum data tersebut. Penyimpangan ini berupa pelanggaran asumsi-asumsi klasik, dapat diketahui dari sisaan model yang terbentuk.

1. Diagnosis Sisaan

Dalam proses pembentukan dan verifikasi model regresi, diagnosis sisaan merupakan tahap yang sangat perlu dilakukan. Tujuannya adalah untuk memeriksa apakah asumsi-asumsi klasik dipenuhi atau tidak, sehingga dapat diketahui apakah dugaan parameternya bersifat BLUE. Selain itu, diagnosis sisaan ini akan dapat memeriksa dan mengidentifikasi apakah ada titik-titik data yang tidak konsisten dengan pola data secara keseluruhan, atau data yang tidak mengikuti pola umum yang telah dimodelkan dengan model regresi.

Metode diagnosis sisaan yang sering digunakan antara lain adalah nilai *Studentized Residual (t-residual)* dan nilai DFITS. Dalam perhitungannya nilai-nilai itu menggunakan acuan vektor-vektor dari matriks HAT. Matriks HAT ini dinyatakan dengan $H = X(X'X)^{-1}X'$. Elemen-elemen diagonalnya merupakan bentuk kuadrat dari

$h_{ii} = x_i (x'x)^{-1}x_i'$, yang hanya tergantung dari nilai-nilai matriks rancangan X .

Suatu titik data dapat diduga sebagai pencilan apabila memiliki nilai t -residual dan atau nilai DFITS yang relatif besar. Menurut Belsley, Kuh & Welsch (1980) dan Ryan (1997) apabila nilai t -residual lebih besar dari 2, atau jika nilai mutlak DFITS yang lebih besar dari

$$2 \left[\frac{p+1}{n} \right]^{1/2},$$

p adalah banyaknya para-

meter, maka titik data itu merupakan pencilan.

2. Regresi Robust Penduga Welsch

Pendugaan parameter menggunakan metode regresi robust berdasarkan pada kasus-kasus yang memiliki nilai t -residual dan atau nilai DFITS yang relatif besar. Pendugaan regresi robust menggunakan perhitungan matematis yang cukup sederhana, tetapi harus dilakukan secara iteratif, sampai diperoleh dugaan terbaik yang memiliki standar error paling kecil.

Banyak ahli statistik yang mengusulkan nilai-nilai penduga pada regresi robust, salah satunya adalah Welsch, sehingga penduganya disebut penduga Welsch. Penduga Welsch digunakan saat asumsi kehomogenan ragam sisaan tidak terpenuhi yang diakibatkan adanya pencilan, sehingga MKT dimodifikasi dengan memberikan suatu fungsi pembobot pada perhitungan-perhitungan pendugaan

parameternya. Metode penyelesaian pendugaan parameter dengan pembobot ini disebut MKT-Terboboti, yang diperoleh dari:

$$\hat{\beta}_w = (X'WX)^{-1} X'WY$$

dimana pembobot $w_i = w(x_i, e_i)$, diselesaikan secara iteratif. Minimalisasi bagian sisaan pada MKT-Terboboti diperoleh dengan meminimumkan bagian sisaan sehingga diharapkan jumlahnya sama dengan nol, yaitu :

$$\sum (y_i - x_i\hat{\beta})w_i x_i = 0$$

Persamaan (2) ini digunakan sebagai dasar perhitungan pada regresi robust penduga Welsch dengan e_i pada w_i diganti dengan e_i/σ sehingga $w_i = w(x_i, e_i/\sigma)$. Nilai σ diduga dengan $s_{(i)}$ yaitu

$$(n-p-2)s_{(i)}^2 = (n-p-2)s^2 - \frac{e_i^2}{1-h_{ii}}$$

dan diduga secara simultan bersama-sama dengan nilai dugaan parameter.

Pembobot yang diusulkan Welsch untuk persamaan (2) adalah

$$w_i = w \left(x_i, \frac{e_i}{s_{(i)}} \right) = \min \left\{ \frac{c}{|d_i|}, 1 \right\}$$

Nilai c adalah konstanta Welsch, d_i adalah nilai DFITS. Pendugaan parameter dengan penduga Welsch dihitung secara iteratif sehingga diperlukan bantuan aplikasi program komputer untuk mempermudah dan mempercepat proses penyelesaian perhitungannya. Salah satu aplikasi program yang dapat digunakan untuk

perhitungan iteratif penduga Welsch adalah Minitab.

METODE PENELITIAN

Penelitian ini akan melakukan pendugaan parameter regresi untuk data amatan yang mengandung pencilan.

1. Metode

Pendugaan parameter dilakukan melalui beberapa tahap, yaitu :

1. Melakukan pendugaan parameter langsung terhadap data amatan menggunakan MKT, dengan anggapan bahwa asumsi klasik terpenuhi.
2. Melakukan diagnosis sisaan terhadap model yang telah terbentuk dengan MKT. Akan dilihat nilai-nilai t-residual dan DFITS untuk setiap titik data amatan. Titik data yang memiliki nilai t-residual dan atau DFITS yang relatif besar (diduga sebagai pencilan) akan dibuang, hal ini disarankan oleh Belsley, Kuh & Welsch (1980), Weisberg (1985) dan Aunuddin (1989), kemudian melakukan pendugaan parameter dengan MKT terhadap sisa data yang ada (MKT tanpa pencilan).
3. Melakukan pendugaan parameter dengan penduga Welsch terhadap data amatan (yang lengkap) tanpa membuang titik data pencilan.

4. Membandingkan hasil 1, 2, dan 3 untuk menentukan model regresi terbaik.

2. Data

Data yang dipergunakan dalam penelitian ini terdiri dari tiga macam data yang mengandung satu, dua dan tiga variabel bebas, yaitu:

1. Data artifisial yang diberikan oleh Anscombe (1977), dalam Aunuddin (1989)
2. Data empiris, yang dikumpulkan oleh Christopher pada *The American Almanac for 1974* dan *The 1974 Word Almanac*. Data ini diambil dari 48 negara bagian di Amerika yang digunakan untuk menyelidiki apakah konsumsi BBM (Bahan Bakar Minyak) merupakan fungsi dari tingkat pajak BBM dan persentase pengemudi yang punya SIM, dalam Juanda (1991).
3. Data empiris dari Brownlee, yang mengungkapkan suatu percobaan di pabrik selama 21 hari percobaan berturut-turut dalam mengoksidasi amonia (NH_3) menjadi asam nitrat HNO_3 , dalam Juanda (1991).

HASIL DAN PEMBAHASAN

Metode regresi robust penduga Welsch akan diterapkan pada masing-masing data amatan. Ketiga macam data amatan itu dianalisis melalui tahapan yang

serupa, sesuai dengan metode yang telah dituliskan di atas.

1. Data Amatan dengan Satu Variabel Bebas

Setelah pendugaan parameter dengan MKT diperoleh, kemudian dilakukan diagnosis sisaan terhadap data amatan. Diagnosis sisaan menghasilkan nilai-nilai t -

residual dan DFITS untuk setiap titik data. Berdasarkan diagnosis sisaan ini, titik data ke-13 dikeluarkan dari amatan karena memiliki nilai t -residual yang lebih besar dari 2 dan nilai DFITS yang besar. Hasil diagnosis sisaan dapat dilihat pada Tabel 1 berikut:

Tabel 1 Diagnosis Sisaan Hasil Pendugaan MKT pada Data dengan Satu Variabel Bebas

| No | Y-duga | Sisaan | t -residual | h_{ii} | DFITS |
|----|--------|---------|---------------|----------|----------|
| 1 | 7.9997 | -0.5397 | -0.44 | 0.1000 | -0.1460 |
| 2 | 7.0003 | -0.2303 | -0.19 | 0.1000 | -0.0620 |
| 3 | 9.4989 | 3.2411 | 1203.54 | 0.2364 | 669.5880 |
| 4 | 7.5000 | -0.3900 | -0.31 | 0.0909 | -0.0990 |
| 5 | 8.4995 | -0.6895 | -0.57 | 0.1273 | -0.2190 |
| 6 | 9.9986 | -1.1586 | -1.16 | 0.3182 | -0.7900 |
| 7 | 6.0008 | 0.0792 | 0.07 | 0.1727 | 0.0300 |
| 8 | 5.0014 | 0.3886 | 0.36 | 0.3182 | 0.2470 |
| 9 | 8.9992 | -0.8492 | -0.74 | 0.1727 | -0.3360 |
| 10 | 6.5006 | -0.0806 | -0.07 | 0.1273 | -0.0250 |
| 11 | 5.5011 | 0.2289 | 0.20 | 0.2364 | 0.1110 |

Perbandingan hasil pendugaan regresi robust penduga Welsch dapat dilihat pada Tabel 2.

Tabel 2. Hasil Pendugaan Koefisien dengan Tiga Metode pada Data dengan Satu Variabel Bebas

| Koefisien regresi | MKT | MKT tanpa pencilan | Welsch |
|-------------------|--------------------|--------------------|--------------------|
| β_0 | 3.0024 (1.124) | 4.006 (0.003) | 4.0044 (0.042) |
| β_1 | 0.4997 (0.1179) | 0.3454 (0.0003) | 0.3456 (0.0046) |

Nilai di dalam tanda kurung merupakan nilai standar deviasi untuk masing-masing koefisien regresi. Terlihat bahwa hasil penduga Welsch memberikan standar deviasi yang relatif sama dengan hasil pendugaan MKT tanpa pencilan. Sedangkan hasil MKT memperlihatkan standar deviasi yang jauh berbeda dengan kedua metode lain.

Hal ini menunjukkan bahwa hasil pendugaan regresi robust penduga Welsch akan memberikan hasil yang sama baiknya dengan MKT saat data pencilannya dibuang terlebih dahulu. Tetapi meskipun demikian, dapat dinyatakan bahwa penduga Welsch memiliki makna yang lebih baik karena dengan pendugaan ini data amatan tidak ada yang dibuang, sehingga tidak ada informasi yang hilang.

2. Data Amatan dengan Dua Variabel Bebas

Melalui metode analisis yang sama, diagnosis sisaan pada data dengan dua variabel bebas menunjukkan adanya dua titik data yang diduga sebagai pencilan yaitu data ke-18 dan data ke-40. Untuk itu, titik data itu dikeluarkan dari analisis terlebih dahulu karena memiliki nilai t-residual dan DFITS yang relatif paling besar, kemudian diterapkan MKT tanpa pencilan untuk mendapatkan koefisien regresi. Terakhir, menggunakan pendugaan dengan penduga Welsch.

Perbandingan hasil pendugaan MKT, MKT tanpa pencilan, dan metode regresi robust penduga Welsch dapat dilihat pada Tabel 3.

Tabel 3. Hasil Pendugaan Koefisien Regresi dengan Tiga Metode terhadap Data dengan Dua Variabel Bebas

| Koefisien regresi | MKT | MKT tanpa pencilan | Welsch |
|-------------------|-----------------|--------------------|-----------------|
| β_0 | 108.971 (171.8) | 198.651(152.4) | 180.584 (160.6) |
| β_1 | -32.075 (12.20) | -30.933 (10.70) | -31.449 (11.19) |
| β_2 | 12.515(2.091) | 10.691 (1.895) | 11.105 (2.003) |

Tabel 3 memperlihatkan bahwa koefisien regresi hasil pendugaan dari ketiga metode relatif berbeda. Akan tetapi, jika diperhatikan nilai standar deviasi untuk koefisien dugaan, terlihat bahwa standar deviasi MKT relatif lebih tinggi dibanding hasil MKT tanpa pencilan dan metode Welsch. Sedangkan standar deviasi hasil penduga Welsch memiliki standar deviasi lebih besar dari hasil MKT tanpa pencilan. Meskipun

demikian, ada informasi yang hilang dari data yang dikeluarkan hasil pendugaan MKT tanpa pencilan, sedangkan hasil penduga Welsch adalah hasil pendugaan tanpa menghilangkan informasi dari data yang ada. Hal ini menunjukkan bahwa hasil pendugaan dengan metode Welsch relatif lebih baik atau lebih cocok dibanding metode klasik (MKT) maupun dengan MKT tanpa pencilan.

3. Data Amatan dengan Tiga Variabel Bebas

Diagnosis sisaan pada data dengan tiga variabel bebas menunjukkan adanya dua titik data yang diduga sebagai pencilan yaitu data ke-4 dan data ke-21. Untuk itu, titik data itu dikeluarkan dari analisis terlebih dahulu karena memiliki nilai t-residual dan DFITS yang relatif paling besar, kemudian

diterapkan MKT tanpa pencilan untuk mendapatkan koefisien regresi. Terakhir, menggunakan pendugaan dengan penduga Welsch.

Perbandingan hasil pendugaan MKT, MKT tanpa pencilan, dan metode regresi robust penduga Welsch dapat dilihat pada Tabel 4

Tabel 4. Hasil Pendugaan Koefisien Regresi dengan Tiga Metode terhadap Data dengan Tiga Variabel Bebas

| Koefisien regresi | MKT | MKT tanpa pencilan | Welsch |
|-------------------|-----------------|--------------------|------------------|
| β_0 | -39.9197(11.90) | -43.704(9.492) | -42.1313 (10.42) |
| β_1 | 0.7156 (0.1349) | 0.8891 (0.1188) | 0.8170 (0.1255) |
| β_2 | 1.2953 (0.3680) | 0.8166 (0.3250) | 1.0156 (0.3430) |
| β_3 | -0.1521(0.1563) | -0.1071 (0.1245) | -0.1258 (0.1368) |

Tabel 4 memperlihatkan bahwa koefisien regresi hasil pendugaan dari ketiga metode relatif tidak jauh berbeda. Akan tetapi jika diperhatikan hasil standar deviasi parameter untuk koefisien dugaan terlihat adanya perbedaan dari ketiga metode. Hasil pendugaan dari MKT tanpa pencilan diperoleh nilai yang lebih kecil dibanding kedua metode yang lain, akan tetapi hasil ini telah menghilangkan informasi yang ada pada data ke-21. Sedangkan hasil dari penduga Welsch diperoleh nilai standar deviasi yang relatif lebih kecil dari MKT, tetapi lebih besar dari MKT tanpa pencilan, akan tetapi besarnya nilai standar deviasi pada hasil ini diimbangi juga dengan naiknya nilai koefisien parameter tersebut, sehingga menghasilkan pengujian yang relatif tidak berbeda dengan

hasil MKT tanpa pencilan. Disamping itu, hasil dari penduga Welsch tidak menghilangkan informasi yang ada pada data.. Hal ini berarti bahwa hasil pendugaan dengan metode robust Welsch relatif jauh lebih baik atau lebih cocok dibanding metode klasik (MKT) maupun MKT tanpa pencilan.

KESIMPULAN DAN SARAN

1. Kesimpulan

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan dapat diambil beberapa kesimpulan sebagai berikut:

1. Dalam analisis regresi linier, hasil penduga MKT dapat digunakan karena lebih mudah dan praktis selama semua asumsi sisaan terpenuhi, akan tetapi

jika terjadi penyimpangan asumsi, maka hasil MKT akan berbias.

2. Jika hasil diagnosis sisaan ternyata menunjukkan adanya pelanggaran asumsi sisaan dari MKT, yang disebabkan adanya pencilan, maka diperlukan tahap analisis data lanjut dan harus membuang data-data pencilannya se-belum menggunakan MKT, akan tetapi hasil ini telah menghilangkan informasi yang ada pada data tersebut.
3. Untuk menghindari adanya penyimpangan asumsi (pencilan data) tanpa menghilangkan informasi yang ada pada data tersebut dapat dilakukan dengan metode regresi robust, salah satu penduga regresi robust yang dapat digunakan adalah penduga Welsch.
4. Hasil dari penduga Welsch, selain menghasilkan dugaan yang lebih baik, juga diperoleh informasi yang utuh dari data yang ada.
5. Dari tiga kasus yang dibahas, diperoleh persamaan regresi terbaiknya adalah hasil dari penduga Welsch:

- a. Model regresi dengan satu variabel bebas adalah:

$$\hat{Y} = 4.004 + 0.3456X$$

- b. Model regresi dengan dua variabel bebas adalah:

$$\hat{Y} = 180.584 - 31.449X_1 + 11.105X_2$$

- c. Model regresi dengan tiga variabel bebas adalah:

$$\hat{Y} = -42.1313 + 0.8170X_1 + 1.0156X_2 - 0.1258X_3$$

2. Saran

Untuk kesempurnaan kajian tentang regresi robust, khususnya penduga Welsch ada beberapa saran yang dapat dilakukan oleh penulis atau peneliti yang tertarik dengan kajian ini, yaitu:

1. Regresi robust adalah analisis yang melibatkan proses iterasi dalam pendugaannya, sehingga disarankan dalam penggunaannya harus bisa mengerti salah satu bahasa program.
2. Untuk dapat menghasilkan keyakinan yang tinggi terhadap efektifitas penduga robust, disarankan untuk melakukan simulasi terhadap beberapa kemungkinan data yang ada.

DAFTAR PUSTAKA

- Aunuddin**, 1989, *Analisis Data*, Bahan Pengajaran, PAU-Ilmu Hayat, IPB, Bogor.
- Belsley, D.A., Kuh, E., & Welsch, R.E.**, 1980, *Regression Diagnostics*. New York: Wiley.
- Draper N.R & H. Smith**, 1992, *Analisis Regresi Terapan*, Gramedia, Jakarta.
- Hadi, A. S., & Simonoff, J.S.**, 1993, Procedures for the Identification of Multiple Outliers in Linear Models, *Journal of the American Statistical Association*, 88, 1264-1272.
- Juanda, B.**, 1991, *Metode Robust Berdasarkan Peringkat dalam Analisis Regresi*, Tesis, Institut Pertanian Bogor. Tidak Dipublikasikan.
- Myers, R. H.**, 1990, *Classical and Modern Regression with Application*, PWS-KENT, Boston.

Neter, J.W., Wasserman & Kutner, M.H., 1990, *Applied Linear Statistical Models, Regression, Analysis of Variance and Experimental Design*. Richard D. Irwin Inc, Illinois.

Roussesuw, P. J., 1984, Least Median of Squares Regression, *Journal of the American Statistical Assosiation*, 79, 871-880.

Ryan, T.P., 1997, *Modern Regression Models*, John Wiley & Sons, New York.

DAFTAR PUSTAKA

Aunuddin, 1989, *Analisis Data Berbasis Pengajaran PAU-timur Hayati* IPB, Bogor.

Bealely, D.A., Kutner, E. & Welsch, R.E., 1980, *Regression Diagnostic*, New York: Wiley.

Dregor M.R. & H. Smith, 1982, *Analisis Regresi Terapan*, Gramedia, Jakarta.

Hadi, A. S. & Simonoff, J.S., 1993, Procedures for the identification of Multiple Outliers in Linear Models, *Journal of the American Statistical Assosiation*, 88, 1284-1272

Juanda, B., 1991, *Metode Robust Berbasis Pengajaran dalam Analisis Regresi*, Tesis Institut Perbanjaan Bogor, Dikembangkan.

Myers, R. H., 1980, *Classical and Modern Regression with Application*, PWS-KENT, Boston.

1. Regresi robust adalah analisis yang melibatkan proses iterasi dalam pengujian hipotesis, sehingga disarankan dalam penggunaannya harus bisa mengidentifikasi satu barisan program.

2. Untuk dapat menghasilkan keyakinan yang tinggi terhadap efektifitas penduga robust, disarankan untuk melakukan simulasi terhadap beberapa kemungkinan data yang ada.

3. Hasil dari penduga Welsch, selain menghasilkan dugaan yang lebih baik, juga diperoleh informasi yang utuh dan data yang ada.

4. Dari tiga kasus yang dibahas, diperoleh gambaran regresi terapkanya adalah hasil dari penduga Welsch:

a. Model regresi dengan satu variabel bebas adalah:

$$\hat{y} = 4.004 + 0.2426X$$

b. Model regresi dengan dua variabel bebas adalah:

$$\hat{y} = 180264 - 31449X_1 + 11103X_2$$

c. Model regresi dengan tiga variabel bebas adalah:

$$\hat{y} = -4213.13 + 0.8170X_1 + 1.0156X_2 - 0.0288X_3$$