

Pengembangan Metode *Weighting by the Absolute Centered External Variable* Pada Data Curah Hujan di Indonesia

Rosanti Hidayati*, Yayat Karyana²

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

*rosantihidayati11@gmail.com, yayatkaryana@gmail.com

Abstract. Indonesia is one of the countries that has a long rainy season. Therefore, the convenience of community activities in Indonesia depends on the weather. Weather prediction is very helpful for community activities. Because a prediction it is necessary to predict the correct model so that it is correct, this research will discuss the development of the WCEV (*Weighting by the Absolute Centered External Variable*) method on heterogeneous variance with DDR (*Diamond Distributed Residuals*) patterns in the January 2005 rainfall data until December 2014 in Indonesia. This method produces weighting for data that has heteroscedasticity patterned DDR. The parameter coefficients can be estimated using the Weighted Least Squares method where the weight applied to each is obtained through the WCEV method. Estimated parameter of the regression coefficient generated by the WCEV method on data with heterogeneous variance with DDR patterns resulted in a smaller MSE (*Mean Square Error*) value compared the MSE value generated by the OLS (*Ordinary Least Square*) method.

Keywords: DDR, Heteroscedasticity, MSE, OLS, WCEV.

Abstrak. Indonesia merupakan salah satu negara yang memiliki durasi musim hujan cukup panjang. Maka dari itu kenyamanan aktifitas masyarakat di Indonesia tergantung terhadap cuaca. Prediksi cuaca sangat membantu berlangsungnya aktifitas masyarakat. Karena dalam sebuah prediksi diperlukan model prediksi yang tepat agar mendekati kebenaran, maka dalam skripsi ini akan dibahas pengembangan Metode WCEV (*Weighting by the Absolute Centered External Variable*) pada varians heterogen dengan pola DDR (*Diamond Distributed Residuals*) pada data data curah hujan bulan Januari tahun 2005 sampai dengan bulan Desember tahun 2014 di Indonesia. Metode ini menghasilkan pembobotan untuk data yang mengalami heteroskedastisitas berpola DDR. Parameter koefisien regresi dapat ditaksir dengan menggunakan Metode Kuadrat Terkecil Terboboti (*Weighted Least Squares*) dimana bobot yang diterapkan pada masing-masing pengamatan diperoleh melalui metode WCEV. Taksiran parameter koefisien regresi yang dihasilkan oleh metode WCEV pada data dengan varians heterogen berpola DDR menghasilkan nilai MSE (*Mean Square Error*) yang lebih kecil dibandingkan dengan nilai MSE yang dihasilkan oleh metode OLS (*Ordinary Least Square*).

Kata Kunci: DDR, Heterogenitas, MSE, OLS, WCEV.

1. Pendahuluan

Dalam melakukan aktifitas keseharian manusia tidak terlepas dari faktor cuaca. Maka dari itu, prediksi cuaca sangat dibutuhkan untuk membantu keberlangsungan aktifitas tersebut. Prediksi bukanlah suatu hal yang pasti dan tepat terjadi, melainkan suatu keadaan yang kejadiannya

mendekati kebenaran dan tidak sepenuhnya sesuai dengan kenyataan. Hasil dari prediksi tersebut akan mendekati kebenaran jika dilakukan dengan metode yang tepat dan benar^[6].

Indonesia merupakan salah satu negara yang memiliki durasi musim hujan cukup panjang. Untuk kenyamanan berlangsungnya aktifitas sehari-hari, kita perlu mengetahui prediksi keadaan cuaca seperti curah hujan. Sesuai dengan siklus hidrologi, curah hujan dipengaruhi beberapa faktor sebagaimana dalam metode regresi sebuah variabel dipengaruhi variabel lainnya untuk mendapatkan sebuah model, dimana model tersebut nantinya bisa dipergunakan sebagai taksiran perkiraan cuaca^[8]. Volume rata-rata curah hujan setiap bulan selalu berbeda dari tahun ke tahun, maka dari itu pada analisis ini curah hujan akan dijadikan sebagai variabel respon dengan bulan sebagai variabel prediktor yang dibuat menjadi variabel dummy karena bulan bersifat kategorik biner.

Dalam model regresi ini ada beberapa asumsi yang harus dipenuhi agar model prediksi yang dibuat menjadi valid sebagai alat prediksi. Sebelum melakukan analisis regresi terdapat beberapa pengujian asumsi yang harus terpenuhi yang disebut uji asumsi regresi yaitu linearitas, normalitas, multikolinearitas, autokorelasi (untuk data deret waktu), dan homogenitas. Apabila seluruh asumsi terpenuhi, maka model regresi linear tersebut dikatakan BLUE (*Best Linear Unbiased Estimator*). Fokus dalam penelitian skripsi ini adalah pada masalah salah satu asumsi yang harus terpenuhi dalam sebuah model regresi yaitu homogenitas varians. Transformasi dilakukan untuk data yang mengalami masalah heteroskedastisitas dengan cara mendeflasi (membagi) masing-masing observasi dibobot (dalam kasus ini dibagi) dengan konstanta baru yang didapat melalui metode baru yang ditemukan oleh Celik (2015) untuk mengoreksi masalah heteroskedastisitas dengan pola bentuk DDR (*Diamond Distributed Residuals*) yaitu metode WCEV (*Weighting by the Absolute Centered External Variable*). Metode WCEV dapat menghasilkan taksiran parameter koefisien regresi yang lebih akurat daripada penduga OLS karena estimasi pada WCEV menghasilkan nilai MSE (*Mean Square Error*) yang lebih kecil dari nilai MSE pada estimasi OLS. Dalam skripsi ini akan ditaksir parameter koefisien regresi pada data curah hujan di Indonesia dengan data curah hujan akan dijadikan sebagai variabel respon dengan bulan sebagai variabel prediktor dengan menggunakan metode WCEV. Berdasarkan uraian diatas, tujuan yang ingin dicapai dari penelitian ini adalah:

1. Apakah pola plot residu pada data curah hujan di Indonesia berbentuk DDR (*Diamond Distributed Residuals*)?
2. Bagaimana taksiran parameter koefisien regresi jika varian heterogen berbentuk DDR (*Diamond Distributed Residuals*) dengan menggunakan WCEV (*Weighting by the Absolute Centered External Variable*)?

2. Landasan Teori

Pembahasan mengenai inferensi statistika standar yang berkaitan dengan regresi linear sederhana maupun regresi linear berganda didasarkan pada metode penaksiran parameter koefisien regresi OLS (*Ordinary Least Square*) yaitu:

$$\hat{\beta} = (X'X)^{-1}X'Y$$

dengan nilai estimasi

$$\hat{Y} = X\hat{\beta}$$

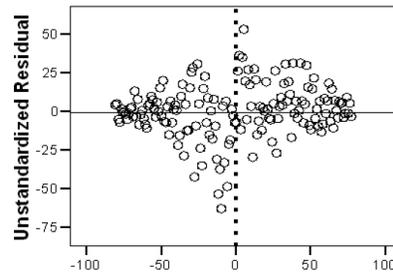
serta nilai yang menunjukkan penyimpangan nilai estimasi terhadap nilai sebenarnya

$$r_i = y_i - \hat{y}_i$$

dimana residu ε_i harus mengikuti distribusi normal dengan varians yang homogen. Apabila asumsi tersebut terpenuhi, maka penaksir OLS ini merupakan penaksir yang baik, artinya bervarians minimum serta hasil dari prediksi akan mendekati kebenaran. Varians yang homogen merupakan asumsi penting pada model regresi linear klasik, artinya semua pengamatan memiliki varians error yang sama^[4]. Jika tidak demikian, yaitu adanya variasi dari pengamatan ke pengamatan maka disitulah terjadi heteroskedastisitas.

Bentuk dari heterogenitas varians bisa membentuk suatu pola tertentu seperti megaphone. Salah satu bentuk dari heteroskedastisitas non-monotonik adalah tipe DDR

(*Diamond Distributed Residuals*) seperti pada berikut ini^[2]:



Gambar 1. tipe DDR (*Diamond Distributed Residuals*)

Perbaikan dilakukan dengan meminimalkan nilai MSE yang diboboti dengan nilai yang memaksimalkan fungsi loglikelihood^[2]. Pembobotan dirumuskan sebagai berikut:

$$\hat{w}_i = \frac{1}{|o_i|^\gamma}, |o_i| \neq 0, \gamma \in [-2, 2]$$

dimana o_i merupakan nilai centered external variable yang diperoleh dari

$$o_i = (d_i - \hat{m}), i = 1, 2, \dots, n$$

serta γ merupakan sebuah skalar yang memaksimalkan fungsi log likelihood dengan rumusan sebagai berikut:

$$L = \frac{1}{2} \left\{ -n \ln(2\pi) - n \ln \sigma^2 + \sum \ln w_i - (n - k) \right\}$$

σ^2 menyatakan nilai MSE yang diperoleh dengan rumusan sebagai berikut:

$$MSE_w = \frac{\sum \hat{W}(Y - \hat{Y})^2}{n - k}$$

sehingga nilai taksiran parameter koefisien regresi baru diperoleh dengan hasil dibawah ini:

$$\hat{\beta}_w = (X' \hat{W} X)^{-1} X' \hat{W} Y$$

Residu yang diperoleh melalui metode OLS dikuadratkan lalu diregresikan kembali seperti persamaan berikut:

$$s_i^2 = |r_i| = \alpha_0 + \alpha_1 d_i + \alpha_2 d_i^2$$

ntuk mendapatkan nilai variabel eksternal dan mengestimasi titik pusat \hat{m} , maka persamaan diatas diturunkan terhadap nilai d dengan hasil sebagai berikut:

$$\frac{\partial s}{\partial d} = \alpha_1 + 2\alpha_2 d$$

sehingga

$$\alpha_1 + 2\alpha_2 d = 0 \rightarrow d = -\frac{\alpha_1}{2\alpha_2}$$

maka taksiran nilai untuk titik pusat \hat{m} berdasarkan nilai d adalah

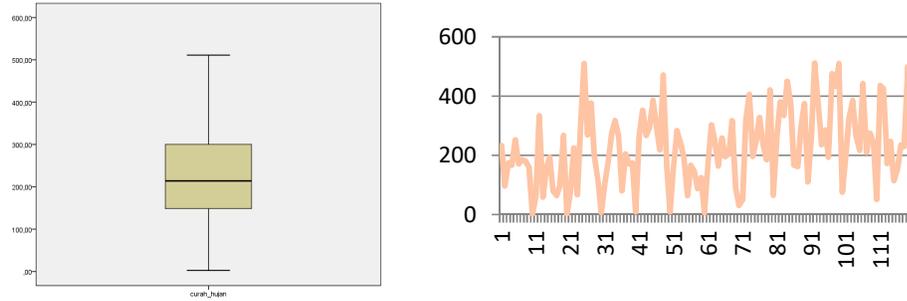
$\hat{m} = -0,5 \frac{\hat{\alpha}_1}{\hat{\alpha}_2}$ Perhitungan untuk nilai MSE disertakan dalam penelitian ini dengan rumusan sebagai berikut:

$$MSE = \frac{\sum \hat{W}(Y - \hat{Y})^2}{n - k}$$

3. Hasil Penelitian dan Pembahasan

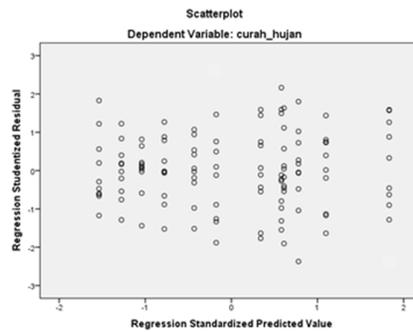
Data yang digunakan dalam analisis pada skripsi ini merupakan data sekunder yang di catat oleh BMKG untuk curah hujan tahun 2005 sampai dengan tahun 2014 dimana data diperoleh berdasarkan Stasiun Klimatologi Deli Serdang. Data yang akan dipakai dalam keperluan

aplikasi pada analisis ini adalah data volume rata-rata curah hujan bulanan di Indonesia tahun 2005 sampai dengan tahun 2014 yaitu sebanyak 120 pengamatan.



Gambar 2. box plot dan dari diagram garis terhadap curah hujan

Berdasarkan hasil dari box plot terhadap curah hujan, terlihat pola pada data mengalami pergeseran rata-rata naik. Berdasarkan hasil dari diagram garis, selama periode bulan Januari 2005 sampai dengan bulan Desember 2014, curah hujan tertinggi berada pada bulan Agustus 2011 dengan volume 511 mm pada data ke-92, sedangkan curah hujan terendah berada pada bulan Agustus 2005 dengan volume 2,4 mm pada data ke-20. Dikarenakan bulan dijadikan sebagai variabel prediktor, maka dibuat variabel dummy karena variabel bulan bersifat kategorik biner.



Gambar 3. Scetterplot curah hujan

Berdasarkan hasil plot residu dengan nilai prediksi seperti pada diatas, menunjukkan pola non-monotonik yang disebutkan oleh Celik (2015) dimana plot membentuk pola tertentu yang disebut dengan pola DDR (*Diamond Distributed Residuals*). Tidak hanya cukup dilihat secara visual saja, untuk menguji apakah dalam model regresi terjadi ketidaksamaan varians dari residual satu pengamatan ke pengamatan yang lain maka akan dilakukan uji formal. Karena data berdistribusi normal maka memenuhi asumsi untuk dilakukan Uji Breusch Pagan dengan hasil sebagai berikut:

Tabel 1. Uji Breusch Pagan

| Breusch-Pagan Test for Heteroskedasticity^{a,b,c} | | |
|--|----|------|
| Chi-Square | df | Sig. |
| 4.807 | 1 | .028 |

a. Dependent variable: curah_hujan

b. Tests the null hypothesis that the variance of the errors does not depend on the values of the independent variables.

c. Predicted values from design: Intercept + jan + feb
+ mar + apr + mei + jun + jul + ags + sep + okt +
nov

Berdasarkan hasil Uji Breusch Pagan, nilai signifikansi menunjukkan kurang dari 0,05 yaitu $0,028 < 0,05$. Maka dari itu, dapat disimpulkan bahwa varians tidak homogen sehingga terdapat heteroskedastisitas varians pada sisaan.

Dengan bantuan software SAS 9.2 diperoleh model regresi $Y = 260,0000 - 98,0000 D.Jan - 127.9600 D.Feb - 45.7600 D.Mar - 82.1500 D.Apr + 12,0000D.Mei - 112.2500 D.Jun - 61.1700 D.Jul - 14.5070D.Ags + 31.2500D.Sep + 75.6500D.Okt + 1.8500D.Nov$. Untuk mendapatkan nilai estimasi optimal (\hat{m}), nilai residual (r_i) diurutkan berdasarkan urutan waktu (dalam kasus ini berdasarkan hari) serta mengabsolutkannya ($|r_i|$). Kemudian membuat variabel eksternal (d) dan variabel eksternal yang dikuadratkan (d^2). Berikutnya variabel eksternal (d) dan variabel eksternal yang dikuadratkan (d^2) diregresikan dengan nilai residual (r_i) yang telah diurutkan dan diabsolutkan ($|r_i|$). Dengan bantuan software SAS 9.2 diperoleh hasil sebagai berikut:

$$|r_i| = 9720.0339 - 11.8021 d + 0.6573 d_2$$

Setelah mendapatkan model residu, kemudian menghitung nilai estimasi optimal (\hat{m}) dengan hasil sebagai berikut:

$$\hat{m} = -0,5 \frac{\hat{\alpha}_1}{\hat{\alpha}_2}$$

$$\hat{m} = -0,5 \frac{-11.8021}{0.6573}$$

$$\hat{m} = 8.9776$$

Nilai dari *centered external variable* (o_i) akan didapatkan berdasarkan nilai estimasi optimal (\hat{m}). Sebelum menghitung estimasi bobot (\hat{w}_i) akan dihitung nilai yang memaksimalkan fungsi loglikelihood dengan bantuan software SAS 9.2 serta diperoleh hasil sebagai berikut:

$$\gamma = 0,2$$

Nilai estimasi bobot (\hat{w}_i) akan didapatkan berdasarkan nilai yang memaksimalkan fungsi loglikelihood. Nilai koefisien regresi baru diperoleh dengan bantuan software SAS 9.2 dengan hasil sebagai berikut:

Tabel 2. Nilai koefisien regresi baru software SAS 9.2

| Variable | Parameter Estimate | MSE | R-Square |
|-----------|--------------------|----------|----------|
| Intercept | 260,22182 | 6575,754 | 0,2422 |
| jan | -102,17985 | | |
| feb | -126,40930 | | |
| mar | -45,91883 | | |
| apr | -81,71449 | | |
| mei | 12,54564 | | |
| jun | -111,74869 | | |
| jul | -60,98013 | | |
| ags | -14,73071 | | |
| sep | 31,38540 | | |
| okt | 75,51809 | | |
| nov | 1,87387 | | |

Berdasarkan hasil diatas, metode WCEV (Weighting by the Absolute Centered External Variable) telah menghasilkan nilai parameter koefisien regresi baru dengan model yaitu $Y = 260.22182 - 102.17985D.jan - 126.40930 D.Feb - 45.91883 D.Mar - 81.71449D.Apr + 12.54564 D.Mei - 111.74869 D.Jun - 60.98013D.Jul - 14.73071D.Ags + 31.38540D.Sep + 75.51809D.Okt + 1.87387D.Nov$.

Parameter koefisien regresi baru yang diperoleh melalui metode WCEV lebih tepat digunakan dari parameter koefisien regresi yang diperoleh berdasarkan metode OLS (Ordinary Least Square) dibawah ini:

Tabel 3. Nilai koefisien regresi baru metode OLS (Ordinary Least Square)

| Variable | Parameter Estimate | MSE | R-Square |
|-----------|--------------------|-------|----------|
| Intercept | 260,00000 | 13556 | 0,2288 |
| jan | -98,00000 | | |
| feb | -127,96000 | | |
| mar | -45,76000 | | |
| apr | -82,15000 | | |
| mei | 12,00000 | | |
| jun | -112,25000 | | |
| jul | -61,17000 | | |
| ags | -14,50700 | | |
| sep | 31,25000 | | |
| okt | 75,65000 | | |
| nov | 1,85000 | | |

Hal tersebut terlihat dari nilai MSE (*Mean Square Error*) yang dihasilkan oleh metode WCEV lebih kecil daripada nilai MSE yang dihasilkan oleh metode OLS. Berikut merupakan tabel perbandingan nilai MSE dengan metode OLS dan nilai MSE dengan metode WCEV:

Tabel 4. perbandingan nilai MSE dengan metode OLS dan nilai MSE dengan metode WCEV

| <i>Mean Square Error</i> | |
|--------------------------|-----------|
| OLS | WCEV |
| 13556 | 6575,7539 |

Dengan menurunnya nilai MSE menggunakan metode WCEV, maka menunjukkan bahwa metode WCEV lebih baik digunakan daripada metode OLS untuk kasus dengan varian heterogen non-monotonik pola DDR karena nilai MSE yang kecil menunjukkan tingkat kesalahan (*error*) yang lebih rendah.

4. Kesimpulan

Berdasarkan penelitian pada data curah hujan bulan Januari tahun 2005 sampai dengan bulan Desember tahun 2014 di Indonesia yang telah dilakukan dan dibahas dalam BAB IV maka dapat disimpulkan bahwa:

1. Data curah hujan memiliki plot residu berpola DDR (Diamond Distributed Residuals).
2. Metode WCEV (Weighting by the Absolute Centered External Variable) pada data dengan residu heterogen non-monotonik pola DDR telah menghasilkan nilai koefisien regresi baru yang lebih efisien daripada nilai koefisien regresi yang diperoleh berdasarkan metode OLS (Ordinary Least Square) karena nilai MSE yang kecil menunjukkan tingkat kesalahan (*error*) yang lebih rendah, sehingga model yang

diperoleh lebih efisien untuk digunakan sebagai alat prediksi curah hujan di Indonesia.

5. Saran

Penelitian ini dapat dikembangkan selanjutnya dengan menambahkan variabel yang belum diteliti untuk menghasilkan pemodelan yang lebih baik serta mengembangkan analisis pada varians heterogen non-monotonik pada kasus lain dengan pola DDR.

Daftar Pustaka

- [1] Carrol, Raymond J, & Ruppert, David. (1988). *Transformation and Weighting in Regression*. New York: Chapman & Hall.
- [2] Çelik, R. (2015). Correcting Diamond Distributed Residuals by WCEV. *Communications in Statistics - Theory and Methods*, 46(19), 9566–9590.
- [3] Çelik, R., & Erar, A. (2016). An Alternative Method Correcting BDR Type of Heteroskedasticity by the Weighting Re-Estimated Absolute Residuals. *Communications in Statistics - Theory and Methods*, 46(23), 11514–11538.
- [4] Gujarati, Damondar N. (2006). *Dasar-Dasar Ekonometrika*. Jakarta: Erlangga.
- [5] Hajarisman, Nusar, & Anneke Iswani Achmad. (2011). *Analisis Regresi Lanjut*. Bandung: Universitas Islam Bandung.
- [6] Irianto, Agus. (2004). *Statistik*. Jakarta: Kencana.
- [7] Margi S, Kristien & Sofian Pandawa W. (2015). Mean Absolute Percentage Error. *Analisa dan Penerapan Metode Single Exponential Smoothing untuk Prediksi Penjualan pada Periode Tertentu*, 260 (Online). (<https://jurnal.umk.ac.id/index.php/SNA/article/view/332>, diakses 5 Juli 2020)
- [8] Soewarno. (2015). *Klimatologi: Pengukuran dan Pengolahan Data Curah Hujan, Contoh Aplikasi Hidrologi dalam Pengelolaan Sumber Daya Air (Seri Hidrologi)*. Yogyakarta: Graha Ilmu.
- [9] Suliadi. (2017). *Bahan Ajar Analisis Regresi*. Bandung: Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Bandung.