



PERBANDINGAN MODEL REGRESI ROBUST DENGAN ESTIMASI *LEAST TRIMMED SQUARE*, *MAXIMUM LIKELIHOOD TYPE*, DAN *SCALE* PADA DATA OUTLIER

JURNAL ILMIAH

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Statistika

Oleh

SUCI WULANDARI

B2A219033

**PROGRAM STUDI S1 STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS MUHAMMADIYAH SEMARANG
2020**

PERSETUJUAN PEMBIMBING

Skripsi dengan judul “Perbandingan Model Regresi Robust dengan Estimasi *Least Trimmed Square*, *Maximum Likelihood Type*, dan *Scale* pada Data **Outlier**” yang disusun oleh:

Nama : Suci Wulandari

NIM : B2A219033

Program Studi : S1 Statistika

Telah disetujui oleh dosen pembimbing pada tanggal : 18 September 2020

Pembimbing Utama


Dr. Rochdi Wasono, M.Si
NIK. 28.6.1026.119

Pembimbing Pendamping


Indah Manfaati Nur, M.Si
NIK. 28.6.1026.221



PENGESAHAN KELULUSAN

Skripsi dengan judul “Perbandingan Model Regresi Robust dengan Estimasi *Least Trimmed Square*, *Maximum Likelihood Type*, dan *Scale* pada Data **Outlier**” yang disusun oleh:

Nama : Suci Wulandari
NIM : B2A219033
Program Studi : S1-Statistika

Telah dipertahankan dalam Sidang Panitia Ujian Skripsi Program Sarjana Universitas Muhammadiyah Semarang Pada Tanggal 18 September 2020.

Panitia Ujian
Ketua Tim Penguji

Tiani Waluya Utami, M. Si
NIK. 28.6.1026.341

Anggota Tim Penguji I



Fathurrokhman Fauzi, M.Stat

Anggota Tim Penguji II



Dr. Rochdi Wasono, M.Si.

NIK. 28.6.1026.119

Anggota Tim Penguji III



Indah Manfaati Nur, M.Si

NIK. 28.6.1026.221

Mengetahui,

Ketua Program Studi


Indah Manfaati Nur, M.Si
NIK. 28.6.1026.221

**SURAT PERNYATAAN
PUBLIKASI KARYA ILMIAH**

Nama : Suci Wulandari
NIM : B2A219033
Fakultas/ Jurusan : S1 Statistika
Jenis Penelitian : Skripsi
Judul : Perbandingan Model Regresi Robust dengan Estimasi *Least Trimmed Square, Maximum Likelihood Type*, dan *Scale* pada Data Outlier
Email : suci_wulan06@yahoo.com

Dengan ini menyatakan bahwa saya menyetujui untuk :

1. Memberikan hak bebas royalti kepada perpustakaan Unimus atas penulisan karya ilmiah saya, demi pengembangan ilmu pengetahuan.
2. Memberikan hak menyimpan, mengalih medikan/mengalih formatkan, mengelola dalam bentuk pengakalan data (*database*), mendistribusikannya, serta menampilkannya dalam bentuk *softcopy* untuk kepentingan akademis kepada perpustakaan Unimus, tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/ pencipta.
3. Bersedia dan menjamin untuk menanggung secara pribadi tanpa melibatkan pihak perpustakaan Unimus, dari semua bentuk tuntutan hukum yang timbul atas pelanggaran hak cipta dalam karya ilmiah ini.

Demikian pernyataan ini saya buat dengan sesungguhnya dan semoga dapat digunakan sebagaimana mestinya.

Semarang, 18 September 2020
Yang menyatakan




Suci Wulandari
NIM. B2A219033

PERBANDINGAN MODEL REGRESI ROBUST DENGAN ESTIMASI *LEAST TRIMMED SQUARE*, *MAXIMUM LIKELIHOOD TYPE*, DAN *SCALE* PADA DATA OUTLIER

Suci Wulandari¹, Dr. Rochdi Wasono, M.Si², Indah Manfaati Nur, M.Si³

¹²³Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Muhammadiyah Semarang

Alamat e-mail : suci_wulan06@yahoo.com

ABSTRAK

Pengangguran merupakan masalah utama yang penanganannya harus terus diupayakan oleh pemerintah. Pengangguran terjadi karena tambahan pencari kerja tidak seimbang dengan kesempatan kerja yang tersedia. Salah satu indikator yang dapat digunakan untuk mengukur pengangguran adalah Tingkat Pengangguran Terbuka (TPT). Provinsi Jawa Timur pada bulan Agustus 2018 memiliki TPT sebesar 3,99%. Angka ini naik dibanding bulan Februari 2018 yang tercatat sebesar 3,85%. Adapun faktor yang mempengaruhi TPT di Jawa Timur adalah Pertumbuhan Ekonomi (PDRB), Indeks Pembangunan Manusia (IPM), dan Tingkat Partisipasi Angkatan Kerja (TPAK). Tujuan dari penelitian ini adalah untuk membandingkan estimasi M, estimasi S, dan estimasi LTS terhadap faktor-faktor yang mempengaruhi TPT di provinsi Jawa Timur Tahun 2018. Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Badan Pusat Statistik Provinsi Jawa Timur. Metode pertama yang digunakan dalam analisis ini adalah Metode Estimasi M. Hasil analisis dari model terbaik diketahui bahwa variabel IPM (X_2) dan TPAK (X_3) adalah variabel yang signifikan. Metode kedua yang digunakan adalah Estimasi S, dimana diketahui variabel yang signifikan adalah PDRB (X_1), IPM (X_2), dan TPAK (X_3). Metode ketiga yang digunakan adalah Estimasi LTS dengan variabel yang signifikan adalah IPM (X_2) dan TPAK (X_3). Berdasarkan hasil analisis didapatkan hasil perbandingan antara ketigametode dimana Estimasi M mempunyai ketepatan sebesar 40,50%, Estimasi S sebesar 22,10% dan Estimasi LTS sebesar 99,99%. Artinya pada kasus ini Estimasi LTS lebih baik dan lebih cocok digunakan dibandingkan Estimasi M dan Estimasi S.

Kata Kunci: Pengangguran, Regresi Robust, Estimasi LTS, Estimasi M, Estimasi S

ABSTRACT

Unemployment is a major problem that the government must strive to deal with. Unemployment occurs because the additional job seekers are not in balance with the available job opportunities. One indicator that can be used to measure unemployment is the Open Unemployment Rate (TPT). East Java Province in August 2018 had TPT of 3.99%. This figure is an increase compared to February 2018 which was recorded at 3.85%. The factors that influence TPT in East Java are Economic Growth (PDRB), Human Development Index (IPM), and Work Force Participation Rate (TPAK). The purpose of this study is to compare the estimation of M, the estimate of S, and the estimate of LTS against the factors that affect the TPT in East Java province in 2018. The data used in this study are secondary data obtained from the Central Statistics Agency of East Java Province. The first method used in this analysis is the M Estimation Method. The results of the analysis of the best models show that the IPM (X_2) and TPAK (X_3) variables are significant variables. The second method used is the S estimate, where it is known that the significant variables are PDRB (X_1), IPM (X_2), and TPAK (X_3). The third method used is LTS estimation with significant variables are IPM (X_2) and TPAK (X_3). Based on the results of the analysis, it was found that the comparison between the three methods where the M estimate was 40.50% accuracy, the S estimate was 22.10% and the LTS estimate was 99.99%. This means that in this case the LTS estimate is better and more suitable to use than the M estimate and the S estimate.

Keywords: Unemployment, Robust Regression, LTS Estimation, M Estimation, S Estimation

Pendahuluan

Indonesia sebagai salah satu negara berkembang mempunyai jumlah penduduk yang cukup padat. Saat ini Indonesia adalah Negara dengan penyumbang jumlah penduduk terbesar keempat di dunia setelah China, India, dan Amerika Serikat[1]. Sebagai negara berkembang, Indonesia dihadapi dengan berbagai permasalahan yang kompleks salah satunya pengangguran. Pengangguran adalah orang yang tidak mempunyai pekerjaan atau tidak mempunyai penghasilan tetapi sedang berusaha mencari pekerjaan[2]. Salah satu indikator yang dapat digunakan untuk mengukur pengangguran adalah Tingkat Pengangguran Terbuka (TPT).

Jawa Timur merupakan salah satu provinsi di Indonesia yang juga dihadapi dengan permasalahan serupa. TPT di Jawa Timur pada bulan Agustus 2018 adalah sebesar 3,99%. Angka ini naik dibanding bulan Februari 2018 yang tercatat sebesar 3,85%[3]. Ini dikarenakan ada beberapa kabupaten/kota di Jawa Timur yang mengalami peningkatan TPT yang cukup signifikan. TPT di Jawa Timur diduga disebabkan oleh beberapa faktor yaitu PDRB, IPM dan TPAK.

Salah satu upaya untuk mengatasi masalah ini adalah dengan menentukan faktor-faktor yang berpengaruh terhadap tingkat pengangguran terbuka di Jawa Timur. Salah satu analisis statistik yang dapat digunakan adalah Analisis Regresi. Dan Metode yang umum digunakan dalam mengestimasi koefisien regresi adalah Metode Kuadrat Terkecil (MKT).

Namun Pada berbagai kasus, tidak jarang ditemukan kondisi dimana asumsi-asumsi pada MKT tersebut tidak terpenuhi. Jika asumsi tidak terpenuhi akan mengakibatkan hasil estimasi parameter pada MKT kurang baik. Hal ini disebabkan karena adanya outlier pada data pengamatan. Oleh Karena itu diperlukan metode lain untuk menangani outlier, metode tersebut adalah metode Regresi Robust. Regresi Robust adalah metode yang digunakan dalam mengatasi outlier tanpa menghapus data outlier tersebut[4]. Metode Regresi robust yang digunakan pada penelitian ini adalah estimasi S, estimasi LTS, dan estimasi M. Estimasi S merupakan salah satu metode yang digunakan untuk meminimumkan standar deviasi sisaan. Estimasi LTS digunakan untuk meminimumkan jumlah kuadrat error. Sedangkan metode estimasi M digunakan untuk meminimumkan fungsi objektif dari residualnya.

Ketiga estimasi tersebut dapat menunjukkan model regresi yang optimal dalam mengetahui faktor-faktor TPT di Jawa Timur Tahun 2018 yang mengandung data outlier.

Landasan Teori

Metode Kuadrat Terkecil (MKT)

Metode Kuadrat Terkecil merupakan salah satu metode untuk mengestimasi parameter pada regresi linear yaitu $\beta_0, \beta_1, \dots, \beta_k$. Tujuan MKT adalah meminimumkan jumlah kuadrat dari kesalahan yang disebut dengan jumlah kuadrat galat terhadap garis regresi.

Terdapat beberapa asumsi yang harus dipenuhi terlebih dahulu sebelum menggunakan regresi linear berganda sebagai alat untuk menganalisis pengaruh variabel-variabel yang diteliti, pengujian asumsi klasik yang digunakan terdiri dari uji autokorelasi, uji heteroskedastisitas, uji multikolinearitas, dan uji normal residual.

Identifikasi Data Outlier

Beberapa metode identifikasi outlier dalam sebuah analisis adalah *scatterplot*, *Leverage Values*, dan *DfFITS* (*Difference fitted Value FITS*). Metode *scatterplot* dilakukan dengan cara memplotkan data dengan observasi ke- i ($i= 1,2, \dots, n$). Selain itu dilakukan dengan cara memplot residual (e) dengan nilai prediksi Y (\hat{Y}). Jika terdapat satu atau beberapa data yang terletak jauh dari pola kumpulan data keseluruhan maka hal ini mengindikasikan adanya outlier. Metode *Leverage Values* digunakan untuk mengukur pengaruh suatu observasi terhadap besarnya estimasi parameter. Untuk beberapa variabel independen, perhitungan nilai *leverage* dapat dilakukan menggunakan persamaan matriks berikut:

$$H = X(X'X)^{-1}X'$$

Pendekatan outlier berdasarkan pada nilai *cutoff* dan apabila nilai h_{ii} melebihi nilai *cutoff* dideteksi sebagai outlier. Sedangkan metode *DfFITS* menampilkan nilai perubahan dalam harga yang diprediksi ketika kasus ke- i dihapuskan dalam penelitian data penelitian yang sudah distandarkan. Perhitungan *DfFITS* adalah sebagai berikut:

$$DfFITS = t_i \left(\frac{h_{ii}}{1-h_{ii}} \right)^{\frac{1}{2}}$$

Data dikatakan outlier apabila nilai $|DfFITS| > 2$

$\sqrt{\frac{p}{n}}$ dengan p adalah banyaknya parameter dan n adalah banyaknya data observasi.

Regresi Robust

Regresi *robust* merupakan alat yang penting untuk menganalisis data yang terkontaminasi oleh *outlier* dan memberikan hasil yang lebih fleksibel. Regresi *robust* tetap menggunakan seluruh data, tetapi dengan memberikan bobot yang kecil untuk data pencilan[5]. Regresi *robust* digunakan untuk mendeteksi pencilan dan memberikan hasil terhadap adanya *outlier*[6]. Terdapat dua hal penting yang sangat diperlukan dalam estimasi regresi *robust* yaitu resistensi dan efisiensi. Suatu estimasi dikatakan resisten terhadap *outlier* jika sebagian kecil dari data tidak dapat memberikan efek yang terlalu besar terhadap estimator sedangkan estimasi memiliki efisiensi yang cukup baik pada berbagai sebaran jika ragamnya mendekati ragam minimum untuk setiap sebaran[7].

Regresi Robust dengan Metode LTS

Estimasi *LTS* merupakan metode penduga regresi *robust* yang menggunakan konsep pengepasan metode kuadrat terkecil untuk meminimumkan jumlah kuadrat sisaan [8]. Penduga *LTS* (β) dinyatakan dalam bentuk rumus sebagai berikut[9]:

$$\hat{\beta}_{LTS} = \arg \min \sum_i^n e_i^2$$

Estimator berdasarkan pada estimasi *S_{LTS}* disebut juga sebagai *Final Weight Scale Estimator* (*FWLS*). Secara matematis fungsi pembobotnya jika nilai $r=3$ sebagai berikut:

$$w_i = \begin{cases} 0 & , \frac{|e_i|}{S_{LTS}} > 3 \\ 1 & , \text{lainnya} \end{cases}$$

Regresi Robust dengan Metode S

Estimasi *S* akan meminimumkan jumlah kuadrat *error* pada persamaan umum regresi linear. Estimasi *S* didefinisikan sebagai berikut :

$$\hat{\beta}_s = \arg \beta \min S_s(e_1, e_2, \dots, e_n)$$

Dimana S_s adalah estimator skala *robust* yang memenuhi $\frac{1}{n} \sum_{i=1}^n \rho \frac{e_i}{\hat{\sigma}_s} = b$ dengan b konstan yang didefinisikan $b = E[z, (\rho)]$, $f(z)$ adalah distribusi normal standar. Nilai *breakdown* dari estimator *S* ditulis $\frac{b}{\max(\rho(e))} = 0,5$ dengan S_s adalah nilai estimator skala *robust* yang minimum dan memenuhi:

$$\min \sum_{i=1}^n \rho \left(\frac{y_i - \sum_{j=0}^k x_{ij} \beta_j}{S_s} \right)$$

dengan: $S_s = \sqrt{\frac{1}{nK} \sum_{i=1}^n w_i e_i^2}$

$K = 0.199$, $w = w_s(u_i) = \frac{\rho(u_i)}{u_i^2}$ dan estimasi awal yang digunakan adalah

$$S_{awal} = \frac{\text{median}|e_i - \text{median}(e_i)|}{0.6745}$$

Estimator $\hat{\beta}$ pada metode regresi *robust* estimasi *S* diperoleh dengan cara melakukan iterasi hingga diperoleh hasil yang konvergen. Proses ini dikenal sebagai *MKT* terboboti secara iterasi yang selanjutnya disebut sebagai *Iteratively Reweighted Least Squared* (*IRLS*) [10].

Regresi Robust dengan Metode M

Metode penaksiran *M* merupakan metode penaksiran dalam regresi *robust* untuk mengestimasi parameter yang disebabkan adanya *outlier*. Penaksiran *M* meminimumkan fungsi ρ (fungsi obyektif) dari residualnya. Fungsi obyektif adalah fungsi yang digunakan untuk mencari fungsi pembobot pada regresi *robust*. Fungsi pembobot yang digunakan adalah fungsi pembobot *Huber* dan fungsi pembobot *Tukey*.

$$\min_{\beta} \rho(e_i) = \min_{\beta} \sum_{i=1}^n \rho(y_i - \sum_{j=0}^k x_{ij} \beta_j)$$

Penaksiran parameter menggunakan metode penaksiran *M* disebut *Iteratively Raweighted Least Square* (*IRLS*). Mengestimasi parameter regresi pada regresi *robust* menggunakan estimasi *M* dilakukan iterasi hingga diperoleh nilai estimasi parameter yang konvergen.

Breakdown Point

Breakdown point adalah persentase terkecil dari banyaknya data yang terkontaminasi atau banyaknya *outlier* yang menyebabkan nilai dari taksiran menjadi besar.. *Breakdown point*

digunakan untuk menjelaskan ukuran ke-robust-an dari teknik *robust*. Kemungkinan tertinggi *breakdown point* untuk sebuah estimator adalah 50%. Jika *breakdown point* lebih dari 50% berarti estimasi model regresi tidak dapat menggambarkan informasi dari kebanyakan data.

Koefisien Determinasi

Koefisien determinasi atau *Adj-Square* merupakan salah satu ukuran yang sederhana dan sering digunakan untuk menguji kualitas suatu persamaan garis regresi[11]. Nilai koefisien determinasi digunakan untuk mengukur seberapa jauh kemampuan model dalam menerangkan variasi variabel dependen. Semakin besar nilai *Adj-Square*, maka semakin besar variasi variabel dependen yang dijelaskan oleh variasi variabel independen. berikut perhitungan \bar{R}^2 :

$$\bar{R}^2 = 1 - \left((1 - R^2) \times \frac{n-1}{n-k} \right)$$

Mean Squared Error (MSE)

Model persamaan yang baik adalah model regresi dengan nilai *MSE* kecil. Semakin kecil nilai *MSE* yang dihasilkan maka semakin baik pendugaan parameter yang dihasilkan tersebut. Nilai *MSE* diperoleh dari nilai jumlah kuadrat galat dibagi dengan *db* jumlah kuadrat sisaan. Berikut perhitungan nilai *MSE*[12]:

$$MSE = \frac{JKG}{n-k}$$

Pengujian Signifikansi Parameter

Uji Overall

Uji *overall* atau uji serentak atau uji *F* merupakan pengujian untuk mengetahui ada atau tidaknya pengaruh secara bersama-sama variabel prediktor terhadap variabel respon. Hipotesis dalam pengujian ini adalah sebagai berikut:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1: \text{Minimal ada satu } \beta_j \neq 0 \text{ untuk } j = 0, 1, \dots, k$$

Dasar pengambilan keputusannya yaitu apabila $F_{hitung} > F_{tabel}$ maka H_0 ditolak. Atau jika $P - value < a$ maka H_0 ditolak.

Uji Parsial

Uji *parsial* atau uji *t* adalah pengujian masing-masing variabel x_i terhadap model. Tujuan dari uji *parsial* ini adalah untuk mengetahui adanya pengaruh antara variabel prediktor ke-*j* dengan $j =$

0, 1, ... *k* dengan variabel respon. Dengan hipotesis

$$H_0: \beta_j = 0$$

$$H_1: \beta_j \neq 0 \text{ untuk } j = 0, 1, \dots, k$$

Pengambilan keputusan statistik uji tersebut adalah apabila $|t_{hitung}| > t_{(1-\frac{\alpha}{2}), n-k-1}$ dengan *k* adalah jumlah parameter, maka tolak H_0 yang artinya terdapat pengaruh variabel *independent* terhadap model.

Fungsi Objektif dan Fungsi Pembobot

Fungsi objektif merupakan representasi pembobot dari residual atau $\rho(u)$, fungsi ini digunakan untuk mencari fungsi pembobot pada regresi *robust* sedangkan fungsi pembobot didapatkan dengan menggunakan fungsi objektif.

Metode Penelitian

Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder Tahun 2018 yang diperoleh dari Badan Pusat Statistik Provinsi Jawa Timur. Data tersebut berjumlah 38 data.

Variabel Penelitian dan Struktur Data

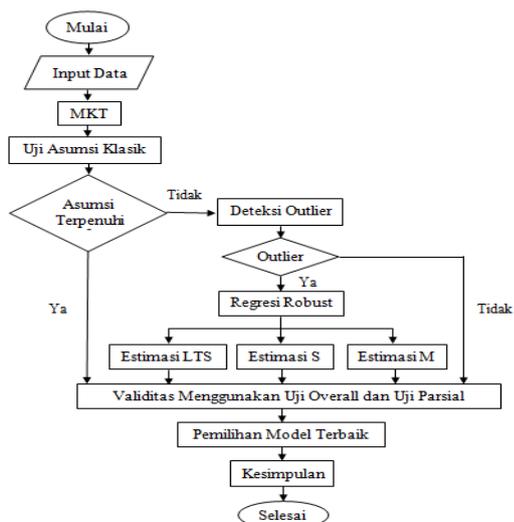
Variabel yang digunakan pada penelitian ini terdiri dari 2 jenis variabel yaitu variabel dependen dan variabel independen. Variabel-variabel tersebut disajikan pada tabel berikut:

Tabel 1. Variabel data

Jenis Variabel	Nama
Y	Tingkat Pengangguran Terbuka
X	X ₁ : Produk Domestik Regional Bruto
	X ₂ : Indeks Pembangunan Manusia
	X ₃ : Partisipasi Angkatan Kerja

Langkah Penelitian

Langkah analisis data dalam analisis dan pembahasan pada Skripsi ini dapat dilihat pada diagram alir di bawah ini.



Gambar 1. Diagram Alir (Flowchart)

Hasil Penelitian dan Pembahasan

Analisis Deskriptif

Analisis deskriptif dalam penelitian ini digunakan untuk melihat gambaran data pada setiap variabel penelitian, berikut deskripsi statistik dari variabel penelitian yang digunakan dalam studi kasus penelitian :

Tabel 2. Analisis Deskriptif Variabel Penelitian

Variabel	Ukuran Statistik		
	Min	Rata – Rata	Maks
TPT	1.43	4.89	9.90
PDRB	0.14	5.08	6.52
IPM	61.94	71.69	82.22
TPAK	63.11	69.48	79.55

Berdasarkan tabel di atas, dapat dilihat bahwa rata – rata tingkat pengangguran terbuka di provinsi Jawa Timur sebesar 4,89% dengan nilai tingkat pengangguran tertinggi sebesar 9,90%. Rata – rata PDRB wilayah Jawa Timur yaitu sebesar 5,08% dengan nilai terendah PDRB yaitu 0,14% dengan rata – rata IPM sebesar 71,69% dan tingkat partisipasi angkatan kerja sebesar 69,48%.

Analisis Regresi Metode Kuadrat Terkecil

Estmasi parameter MKT dengan menggunakan software R diperoleh sebagai berikut:

Tabel 3. Hasil Estimasi Parameter

Parameter	Nilai Estimasi	\bar{R}^2
Intercept	1,943	
PDRB (X_1)	0,151	0,209
IPM (X_2)	0,196	
TKPT (X_3)	-0,170	

Sehingga diperoleh model sebagai berikut:

$$\hat{Y} = 1,943 + 0,151X_1 + 0,196X_2 - 0,170X_3$$

diperoleh \bar{R}^2 sebesar 0,209 yang artinya bahwa variabel dependen (Y) mampu dijelaskan oleh X_1 , X_2 , dan X_3 sebesar 20,9% sedangkan sisanya dijelaskan oleh variabel lain. Model regresi yang diperoleh dari metode kuadrat terkecil dikatakan menghasilkan estimator yang tiak bias atau BLUE ketika uji asumsi klasik dapat terpenuhi terlebih dahulu. Berikut adalah hasil uji asumsi klasik yang diperoleh :

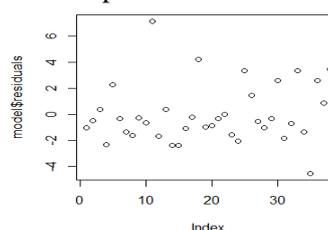
Tabel 4. Uji Asumsi Klasik

Uji Asumsi	Nilai	Kesimpulan
Autokorelasi	DW (2,2161)	Terpenuhi
Multikolinieritas	VIF (< 10)	Terpenuhi
Normalitas residual	Pvalue (0,000568)	Tidak Terpenuhi
Heterokedastisitas	Pvalue (0,547)	Terpenuhi

Berdasarkan keempat asumsi di atas, dapat dilihat bahwa asumsi asumsi normalitas residual tidak terpenuhi. Karena ada asumsi yang tidak terpenuhi maka akan mengakibatkan hasil estimasi parameter pada MKT kurang baik. Hal ini disebabkan adanya outlier pada data pengamatan. Sehingga perlu dilakukan identifikasi outlier.

Identifikasi Outlier

Pada uji asumsi klasik sebelumnya didapatkan kesimpulan bahwa data tidak berdistribusi normal, hal ini dapat diduga karena adanya outlier pada data tersebut. Oleh karena itu perlu dilakukannya identifikasi outlier. Berikut hasil plot residual pada data:



Gambar 2. Plot Residual pada Data

Pada gambar di atas dapat dilihat bahwa terdapat beberapa titik yang memencil dari sebaran data lainnya, titik tersebut disebut dengan *outlier*. Namun, plot tersebut tidak dapat memberikan informasi data mana yang termasuk kedalam *outlier*, oleh karena itu untuk mengidentifikasi *outlier* selain menggunakan *scatter plot* digunakan metode lain yaitu metode *leverage* yang didasarkan pada nilai $cuttoff = \frac{2p}{n} = \frac{2.4}{38} = 0,210$ kemudian nilai h_{ii} yang didapatkan berdasarkan perhitungan dan dibandingkan dengan nilai *cuttoff* yang ada. Dengan daerah kritis $h_{ii} > cuttoff$ maka dideteksi *outlier*.

Tabel 5. Hasil Uji Leverage

No	h_{ii}	Keputusan
1	0,2499	> 0,210 Data <i>Outlier</i>
26	0,3693	
27	0,2647	
29	0,5578	

Berdasarkan tabel di atas hasil identifikasi *outlier* dengan menggunakan persamaan *leverage* maka data yang mempunyai nilai lebih dari *cuttoff* adalah data ke-1, 26, 27, dan 29 yang merupakan data outlier pada arah x.

Selain menggunakan metode *leverage* terdapat metode *DfFITS* dimana suatu data dikatakan *outlier* ketika $|DfFITS| > 0,648$

Tabel 6. Hasil Nilai DfFITS

No	Nilai DfFITS	Keputusan
11	1,568	> 0,648 Data <i>Outlier</i>
35	0,697	

Berdasarkan tabel diatas didapatkan beberapa data memiliki nilai lebih dari 0,648 yaitu pada data observasi ke-11 dan 35. Sehingga dapat disimpulkan bahwa data observasi tersebut *outlier* pada X dan Y.

Analisis Regresi *Robust* Estimasi M

Berdasarkan *output* dihasilkan estimasi parameter dengan iterasi 0 merupakan hasil estimasi MKT. Hasil estimasi parameter regresi *robust* hingga iterasi 9 kemudian pada iterasi 8 estimasi parameter regresi *robust* estimasi M sudah dianggap konvergen yang dibuktikan oleh tabel berikut:

Tabel 7. Hasil Iterasi Estimasi M

Iterasi	Estimasi M			
	β_0	β_1	β_2	β_3
0	1,943	0,151	0,196	-0,170

1	3,048	0,026	0,239	-0,224
2	3,127	-0,022	0,250	-0,235
3	3,094	-0,035	0,252	-0,236
4	3,121	-0,034	0,251	-0,235
5	3,118	-0,034	0,251	-0,235
6	3,116	-0,034	0,251	-0,235
7	3,115	-0,034	0,250	-0,235
8	3,114	-0,034	0,250	-0,235
9	3,114	-0,034	0,250	-0,235

Dari tabel di atas didapatkan estimasi parameter terbaik menggunakan pembobot Huber terletak pada iterasi ke-8. Sehingga dibentuk model persamaan regresi *robust* estimasi M Huber sebagai berikut.

$$\hat{Y} = 3,114 - 0,034X_1 + 0,250X_2 - 0,235X_3$$

Selanjutnya, model persamaan di atas dilakukan pengujian parameter menggunakan uji *overall* yang merupakan pengujian serentak semua parameter dalam model regresi.

Tabel 8. Validitas Estimasi M

Parameter	Nilai Estimate	$ t_{hitung} $	t_{tabel}	Keputusan	P-Value	Keputusan
Intercept	3,114	0,330	2,032	Terima H0	0,000	Tolak H0
X ₁	-0,034	0,113		Terima H0		
X ₂	0,250	3,127		Tolak H0		
X ₃	-0,235	2,424		Tolak H0		

Berdasarkan hasil pengujian yang terdapat pada Table 4.10 di atas, disimpulkan bahwa tolak H0 karena $p_value > \alpha$ yaitu $0,000 > 0,05$. Artinya model layak digunakan. Uji validitas selanjutnya adalah uji parsial untuk mengetahui adanya pengaruh variabel independen terhadap variabel dependen.

Daerah kritis dari uji parsial ini adalah $|t_{hitung}| > t_{tabel}$ maka tolak H0. Tabel di atas menunjukkan bahwa variabel X₂ dan X₃ secara signifikan mempengaruhi variabel dependen sehingga model regresi M yang terbentuk adalah:

$$\hat{Y} = 3,114 + 0,250X_2 - 0,235X_3$$

Selanjutnya, dilakukan perhitungan nilai \bar{R}^2 dan MSE sehingga didapatkan hasil sebagai berikut.

Tabel 9. Nilai \bar{R}^2 dan MSE Estimasi M

Metode	\bar{R}^2	MSE
Estimasi M	0,405	3,010

Berdasarkan tabel di atas diperoleh nilai MSE sebesar 3,010 karena MSE adalah rata-rata residual dikuadratkan maka artinya rata-rata kesalahan dalam memprediksi \hat{Y} sebesar 1,735. Sedangkan metode estimasi M mempunyai

\bar{R}^2 sebesar 0,405 artinya 40,5 % variasi Y dapat dijelaskan oleh variabel independen sedangkan sisanya dijelaskan oleh variabel lain.

Analisis Regresi Robust Estimasi S

Berdasarkan *output* dihasilkan estimasi parameter dengan iterasi 0 merupakan hasil estimasi MKT. Hasil estimasi parameter regresi *robust* hingga iterasi 14 kemudian pada iterasi 13 estimasi parameter regresi *robust* estimasi S sudah dianggap konvergen yang dibuktikan oleh tabel berikut:

Tabel 10. Hasil Iterasi Estimasi S

Iterasi	Estimasi S			
	β_0	β_1	β_2	β_3
0	1,943	0,151	0,196	-0,170
1	2,837	-0,007	0,239	-0,221
2	3,162	0,102	0,214	-0,204
3	2,437	0,131	0,203	-0,184
4	2,254	0,138	0,200	-0,179
5	2,206	0,140	0,200	-0,178
⋮	⋮	⋮	⋮	⋮
12	2,190	0,141	0,199	-0,177
13	2,190	0,141	0,199	-0,177
14	2,190	0,141	0,199	-0,177

Dari tabel di atas, dapat dibentuk model persamaan regresi *robust* estimasi S sebagai berikut:

$$\hat{Y} = 2,190 + 0,141X_1 + 0,199X_2 - 0,177X_3$$

Seperti pada langkah estimasi sebelumnya, model persamaan di atas perlu dilakukan validasi uji *overall*

Tabel 11. Validitas Estimasi S

Parameter	Nilai Estimate	$ t_{hitung} $	t_{tabel}	Keputusan	P-Value	Keputusan
Intercept	2,190	1,079		Terima H0		
X ₁	0,141	2,178	2,032	Tolak H0	0,012	Tolak H0
X ₂	0,199	11,568		Tolak H0		
X ₃	-0,177	-8,509		Tolak H0		

Berdasarkan hasil pengujian yang terdapat pada Table 4.13 di atas, disimpulkan bahwa tolak H0 karena $p_value > \alpha$ yaitu $0,012 > 0,05$ maka model layak digunakan. Selanjutnya dilakukan uji parsial dengan daerah kritis yang digunakan dalam uji parsial ini adalah $|t_{hitung}| > t_{tabel}$ maka tolak H0. Tabel di atas menunjukkan bahwa variabel X₁, X₂, dan X₃ secara signifikan mempengaruhi variabel dependen Y karena masing-masing nilai $|t_{hitung}|$ variabel tersebut $> t_{tabel}$ sehingga model regresi estimasi S yang terbentuk adalah:

$$\hat{Y} = 2,19 + 0,141X_1 + 0,199X_2 - 0,177X_3$$

Selanjutnya, dilakukan perhitungan nilai \bar{R}^2 dan MSE sehingga didapatkan hasil sebagai berikut.

Tabel 12. Nilai \bar{R}^2 dan MSE Estimasi S

Metode	\bar{R}^2	MSE
Estimasi S	0,221	5,439

Berdasarkan tabel di atas diperoleh nilai MSE sebesar 5,439 karena MSE adalah rata-rata residual dikuadratkan maka artinya rata-rata kesalahan dalam memprediksi \hat{Y} sebesar 2,332. Sedangkan metode estimasi S mempunyai \bar{R}^2 sebesar 0,221 artinya 22,1% variasi Y dapat dijelaskan oleh variabel independen sedangkan sisanya dijelaskan oleh variabel lain.

Analisis Regresi Robust Estimasi LTS

Berdasarkan *output* dihasilkan estimasi parameter dengan iterasi 0 merupakan hasil estimasi MKT. Hasil estimasi parameter regresi *robust* hingga iterasi 7 kemudian pada iterasi 6 estimasi parameter regresi *robust* estimasi LTS sudah dianggap konvergen yang dibuktikan oleh tabel berikut:

Tabel 13. Hasil Iterasi Estimasi LTS

Iterasi	Estimasi LTS			
	β_0	β_1	β_2	β_3
0	1,943	0,151	0,196	-0,170
1	5,016	0,007	0,206	-0,221
2	3,270	0,055	0,196	-0,192
3	7,942	0,074	0,165	-0,227
4	8,513	0,079	0,161	-0,232
5	8,218	0,080	0,163	-0,230
6	8,135	0,078	0,164	-0,230
7	8,135	0,078	0,164	-0,230

Dari tabel di atas, dapat dibentuk model persamaan regresi *robust* estimasi S sebagai berikut:

$$\hat{Y} = 8,135 + 0,078X_1 + 0,164X_2 - 0,230X_3$$

Seperti pada langkah estimasi sebelumnya, model persamaan di atas perlu dilakukan validasi uji *overall*

Tabel 14. Validitas Estimasi LTS

Parameter	Nilai Estimate	$ t_{hitung} $	t_{tabel}	Keputusan	P-Value	Keputusan
Intercept	8,135	1,569		Terima H0		
X ₁	0,078	0,790	2,032	Terima H0	0,036	Tolak H0
X ₂	0,164	3,934		Tolak H0		
X ₃	-0,230	5,238		Tolak H0		

Berdasarkan hasil pengujian yang terdapat pada table di atas, disimpulkan bahwa tolak H0 karena $p_value > \alpha$ yaitu $0,036 > 0,05$ maka model layak digunakan. Selanjutnya dilakukan uji parsial dengan daerah kritis yang digunakan dalam uji parsial ini adalah $|t_{hitung}| > t_{tabel}$ maka tolak H0. Tabel di atas menunjukkan bahwa variabel X₂ dan X₃ secara signifikan mempengaruhi variabel dependen Y. Karena masing-masing nilai $|t_{hitung}|$ variabel tersebut $> t_{tabel}$ sehingga model regresi estimasi LTS yang terbentuk adalah:

$$\hat{Y} = 8,135 + 0,164X_2 - 0,230X_3$$

Selanjutnya, dilakukan perhitungan nilai \bar{R}^2 dan MSE sehingga didapatkan hasil sebagai berikut.

Tabel 15. Nilai \bar{R}^2 dan MSE Estimasi LTS

Metode	\bar{R}^2	MSE
Estimasi LTS	0,999	0,005

Pemilihan Metode Estimasi Terbaik

Berikut hasil perbandingan dari masing-masing estimasi pada regresi *robust* menggunakan nilai \bar{R}^2 dan MSE:

Tabel 16. Perbandingan Nilai \bar{R}^2 dan MSE

Parameter	Estimate S	Estimate LTS	Estimate M
Intercept	2,190	8,135	3,114
X ₁	0,141	0,078	-0,034
X ₂	0,199	0,164	0,250
X ₃	-0,177	-0,230	-0,235
\bar{R}^2	0,221	0,999	0,405
MSE	5,439	0,005	3,010

Metode terbaik adalah metode yang memiliki MSE terkecil serta \bar{R}^2 paling besar diantara estimasi lain. Berdasarkan Tabel 4.18 dapat dilihat bahwa nilai \bar{R}^2 paling tinggi dimiliki oleh estimasi LTS sedangkan nilai MSE terkecil juga dimiliki oleh estimasi LTS, maka dapat disimpulkan bahwa metode estimasi LTS merupakan metode estimasi *robust* yang paling baik digunakan dalam mengestimasi parameter regresi untuk data tingkat pengangguran terbuka di Jawa Timur Tahun 2018.

Interpretasi Model

Berdasarkan proses validasi model yaitu uji parsial maka terbentuk model persamaan regresi *robust* estimasi LTS sebagai berikut:

$$\hat{Y} = 8,135 + 0,164X_2 - 0,230X_3$$

Model di atas mempunyai nilai \bar{R}^2 sebesar 0,999 artinya 99,9% variasi variabel dependen dapat dijelaskan oleh variabel independen dan sisanya dipengaruhi oleh faktor lain.

Koefisien regresi variabel IPM (X₂) sebesar 0,164 artinya dengan menganggap variabel independen lain konstan, setiap perubahan 1 satuan IPM maka tingkat pengangguran terbuka akan mengalami perubahan sebesar 0,164. Sedangkan koefisien regresi variabel TPAK (X₃) sebesar -0,230 artinya dengan menganggap variabel independen lain konstan, setiap perubahan 1 satuan TPAK maka tingkat pengangguran terbuka akan mengalami perubahan sebesar 0,230.

Kesimpulan

Faktor-faktor yang mempengaruhi tingkat pengangguran terbuka di Jawa Timur Tahun 2018 adalah indeks pembangunan manusia dan partisipasi angkatan kerja. Dan metode estimasi *robust* yang paling baik digunakan pada data tingkat pengangguran terbuka di Jawa Timur Tahun 2018 adalah estimasi LTS karena estimasi LTS memiliki nilai \bar{R}^2 terbesar dan MSE terkecil dibandingkan metode estimasi lainnya. Sedangkan Model terbaik tingkat pengangguran terbuka di Jawa Timur Tahun 2018 dengan menggunakan metode estimasi LTS adalah sebagai berikut:

$$\hat{Y} = 8,135 + 0,164X_2 - 0,230X_3$$

Daftar Pustaka

- [1] World Bank. (2018). Jumlah Penduduk Indonesia Tahun 2018. Diambil pada [Biro Sensus Amerika Serikat](#), [Bank Dunia](#)
- [2] Murni, A. 2006. *Ekonomika Makro*. Bandung: Refika Aditama.
- [3] Badan Pusat Statistik. 2018. Tingkat Pengangguran Terbuka Kabupaten/ Kota di Jawa Timur 2018. Diambil pada <https://jatim.bps.go.id/statictable>
- [4] Mashitah, Wibowo, A., & Indriani, D. (2013). Metode Robust Regression on Ordered

Statistics (ROS) pada Data Tersensor Kiri dengan Outlier. *Jurnal Biometrika dan Kependudukan*, Vol. 2, No. 2 , 148–157.

- [5] Soemartini, 2007. *Pencilan (Outlier)*. Bandung: Universitas Padjajaran.
- [6] Chen, C. 2002. Robust Regression and Outlier Detection with the Robustreg procedure. SUGI Proceedings. SAS institute Inc., Cary, NC.:265-27.
- [7] Montgomery, D. C., Peck, E. A., dan Vining, G. G. 1982. *Intoducing to Linear Regression analysis*. New York: John Whilley and Sons Inc.
- [8] Akbar, dan Maftukhah. 2007. Optimasi Kekuatan *Tourqe* pada Lampu TL. *Jurnal ilmiah Sains dan Teknologi*. 6(1): 218-229.
- [9] Rousseuw, P. J., dan Leroy, A.M. 1987. *Robust Regression and Outlier Detection*. New York: John Wiley dan Sons.
- [10] Fox, J. dan Weisberg, S. 2010. *Robust Regression in R. Apendix to an R and S-Plus Companion to Applied Regression. Second Edition*.
- [11] Gujarati. 1978. *Basic Econometrics. 1st Edition*. Newyork: McGraw-Hill. Terjemahan Zain, S.1988. *Ekonometrika Dasar*. Erlangga.
- [12] Sembiring, R.K.2003. *Analisis Regresi*, Edisi Kedua. Bandung: ITB

