REGRESI KUANTIL PADA DATA YANG MENGANDUNG PENCILAN

(Skripsi)

Oleh

Annisa Rizki Utami

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2018

ABSTRACT

QUANTILE REGRESSION TO DATA WHICH CONTAIN OUTLIERS

By

ANNISA RIZKI UTAMI

The purpose of this research is to find out the accuracy of the quantile regression method in estimating models on stimulation data which contain outliers and to compare value with ordinary least square. The result shows that the method performs best to data contain outliers less than 50% of the total amount of data and quantile regression produces a better coefficients estimation than ordinary least square.

Key words: Regression, OLS, Outliers, Quantile Regression.

ABSTRAK

REGRESI KUANTIL PADA DATA YANG MENGANDUNG PENCILAN

Oleh

ANNISA RIZKI UTAMI

Tujuan dari penelitian ini adalah untuk mengetahui keakuratan metode regresi kuantil dalam menduga model pada data simulasi yang mengandung pencilan dan membandingkan nilai dugaannya dengan metode kuadrat terkecil. Hasil menunjukkan bahwa metode egresi kuantil baik digunakan pada data dengan kontaminasi pencilan kurang dari sama dengan 50% dari jumlah data dan regresi kuantil menghasilkan dugaan koefisien regresi yang lebih baik daripada metode kuadrat terkecil.

Kata Kunci: Regresi, MKT, Pencilan, Regresi Kuantil.

REGRESI KUANTIL PADA DATA YANG MENGANDUNG PENCILAN

Oleh

ANNISA RIZKI UTAMI

Skripsi

Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA SAINS

pada

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2018

Judul Skripsi : **REGRESI KUANTIL PADA DATA YANG MENGANDUNG PENCILAN**

Nama Mahasiswa : *Annisa Rizki Utami*

Nomor Pokok Mahasiswa : 1417031018

Program Studi : Matematika

Fakultas : Matematika dan Ilmu Pengetahuan Alam

MENYETUJUI

1. Komisi Pembimbing

Ir. Netti Herawati, M.Sc., Ph.D.	**Nusyirwan, Drs., M.Si.**
NIP 19650125 199003 2 001	NIP 19661010 199205 1 001

2. Ketua Jurusan Matematika

Prof. Dra. Wamiliana, M.A., Ph.D.
NIP 19631108 198902 2 001

MENGESAHKAN

1. Tim Penguji

Ketua : **Ir. Netti Herawati, M.Sc., Ph.D.** ....

Sekretaris : **Nusyirwan, Drs., M.Si.** ....

Penguji
Bukan Pembimbing : **Rudi Ruswandi, Drs., M.Si.** ....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Prof. Warsito, S.Si., D.E.A., Ph.D.
NIP 19710212 199512 1 001

Tanggal Lulus Ujian Skripsi : **26 Juli 2018**

PERNYATAAN SKRIPSI MAHASISWA

Yang bertanda tangan di bawah ini :

Nama	: **Annisa Rizki Utami**
Nomor Pokok Mahasiswa	: **1417031018**
Jurusan	: **Matematika**
Judul Skripsi	: **Regresi Kuantil Pada Data Yang Mengandung Pencilan**

Dengan ini menyatakan bahwa penelitian ini adalah hasil pekerjaan saya sendiri. Dan Apabila kemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar lampung, Juli 2018

Yang Menyatakan,

METERAI TEMPEL
D9E9CAFF05359
6000
ENAM RIBU RUPIAH

Annisa Rizki Utami

RIWAYAT HIDUP

Penulis bernama lengkap Annisa Rizki Utami, dilahirkan pada tanggal 14 September 1996 di Lampung Timur. Penulis merupakan putri sulung dari Bapak Baihaki dan Ibu Tuti Herawati, serta kakak dari Novika Justicia dan Athaya Ghina Aprilia.

Penulis menempuh pendidikan di TK IT Baitul Muslim pada tahun 2000 sampai 2002, SD IT Baitul Muslim pada tahun 2002 sampai 2008. Kemudian melanjutkan ke sekolah menengah pertama di SMP IT Baitul Muslim pada tahun 2008 sampai 2011. Dan belajar pada jenjang SMA di SMA Negeri 1 Way Jepara pada tahun 2011 sampai 2014.

Pada tahun 2014, melalui jalur SBMPTN, penulis diterima dan terdaftar sebagai mahasiswa S1 Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

Pada tahun 2014 penulis terdaftar sebagai anggota magang Bidang Eksternal Himatika Unila, ESO, serta ROIS dan pada tahun 2015-2016 terdaftar sebagai anggota Departemen Pengembangan Sumber Daya Mahasiswa (PSDM) BEM

FMIPA Unila. Pada periode 2016, penulis dipercaya menjadi Sekretaris Bidang Keilmuan Himatika Unila.

Di awal tahun 2017 penulis melakukan Kerja Praktik (KP) di Kantor Badan Pusat Statistik Lampung Timur. Di pertengahan tahun 2017, sebagai bentuk aplikasi bidang ilmu kepada masyarakat, penulis telah melaksanakan Kuliah Kerja Nyata Kebangsaan (KKNK) selama 40 hari di Desa Taludaa, Kecamatan Bone, Kabupaten Bone Bolango, Provinsi Gorontalo.

KATA INSPIRASI

"Diwajibkan atas kamu berperang, padahal itu tidak menyenangkan bagimu. Tetapi boleh jadi kamu tidak menyenangi sesuatu, padahal itu baik bagimu. Dan boleh jadi kamu menyukai sesuatu, padahal itu tidak baik bagimu. Allah mengetahui, sedang kkamu tidak mengetahui"

(QS. Ar-Baqarah: 216)

"Hai orang-orang yanga beriman, jika kamu (menolong) agama Allah, niscaya Dia akan menolongmu dan meneguhkan kedudukanmu."

(Q.S. Muhammad: 7)

"Tuhan hanya memberi yang terbaik, tetapi diri kita tidak jarang hanya menginginkan yang sepertinya terlihat baik."

(Falafu)

"Impianmu, lekas digapai. Jika belum sampai, kurangi bersantai. Semoga tercapai. Sebelum semangat tercerai berai dan hati sibuk berandai-andai."

(Randiaputra)

Do good and good will come to you.

(Annisa Rizki Utami)

PERSEMBAHAN

Alhamdulillahirobbil'alamin,
Puji dan syukur kita panjatkan kepada Allah Subhanahu Wata'ala karena atas berkah dan nikmat-Nya kepada kita, Shalawat serta salam selalu tercurah kepada Nabi Muhammad Shallallahu 'Alaihi Wasallam yang telah memberikan kabar gembira kepada umat manusia.

Kupersembahkan karya yang sangat sederhana ini untuk:

Papa dan Mama
Tidak ada kata yang dapat Sejati sampaikan untuk kalian kecuali terimakasih yang sebesar-besarnya atas semua yang telah kalian berikan untukku. Cinta, kasih sayang, waktu, pengorbanan, dan keringat yang belum bisa aku balas. Terimakasih karena selalu mendoakan dan mendukung setiap langkah yang sejati pilih. Karena ridho Allah berawal dari ridho kalian.

Adikku Vika dan Ghina
Terimakasih telah mengajarkan sejati banyak hal, terutama arti kesabaran. Doakan sejati agar bisa menjadi sosok kakak yang lebih baik lagi.

Sahabat-sahabatku.

SANWACANA

Alhamdulillah, puji dan syukur penulis panjatkan kepada Allah SWT karena atas limpahan karunia serta ridho-Nya sehingga skripsi dengan judul **“Regresi Kuantil Pada Data Yang Mengandung Pencilan”** dapat terselesaikan.Shalawatserta salam selalu tercurahkan kepada suri tauladan kita Nabi Muhammad SAW. Dalam menyelesaikan skripsi ini, penulis menyadari banyaknya bimbingan, bantuan, dan dukungan berbagai pihak. Untuk itu penulis mengucapkan terimakasih kepada :

1. Ibu Ir. Netti Herawati, M.Sc., Ph.D., selaku pembimbing I yang senantiasa membimbing dan memotivasi penulis dalam menyelesaikan skripsi ini.
2. Bapak Nusyirwan, Drs., M.Si., selaku pembimbing II yang selalu memberikan dukungan dan arahan kepada penulis.
3. Bapak Rudi Ruswandi, Drs., M.Si., selaku penguji yang telah memberikan saran dan semangat sehingga terselesainya skripsi ini.
4. Bapak Agus Sutrisno, S.Si., M.Si., selaku Pembimbing Akademik yang telah membimbing penulis.
5. Ibu Prof. Dra. Wamiliana, M.A, Ph.D., selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam.
6. Bapak Prof. Warsito, S.Si., D.E.A., Ph.D., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam.

7. Seluruh Dosen, staf dan karyawan Jurusan Matematika FMIPA Universitas Lampung.
8. Papa, Mama, Uti Vika, dan Adek Ghina yang tidak pernah lelah mendo'akan, mendukung dan memberi perhatian kepada penulis.
9. Keluarga Besar H. Maulana Abbas dan H. Mukti, serta Keluarga Besar M. Zahari Mustafa yang telah banyak memberi semangat dan motivasi untuk membantu penulis.
10. Teman-teman seperjuangan Lalak, Amanda, Rara, Ecak, Pau, Suci, Dhea, Vivi, Kodir, Aldi, dan Ijul yang telah banyak membantu penulis.
11. Sahabat-sahabat fii sabilillahku Oca, Ninda, Chintia, Lutfi, dan Deni yang tidak pernah lelah memberikan semangat kepada penulis.
12. Sahabat-sahabat kontrakan tercintaku Clara, Eyi, Astika, Dina, dan Ekaw.
13. Teman-Teman KKN Kebangsaanku Neisya, Hasni, Lisa, Alya, Cut, Sirman, Hendra, dan Firman yang selalu mendukung dan mendoakan.
14. Teman-temanku dan Keilmuan 2016 yang selalu memotivasi penulis.
15. Teman-teman satu bimbingan Mona, Uti, Uung, Jelli, Fietra, Abdul, Aldo, dan Alvin yang telah banyak membantu.
16. Teman-teman Matematika 2014 yang telah memberikan pengalaman luar biasa.
17. Keluarga besar Bidang Keilmuan dan HIMATIKA FMIPA UNILA.
18. Dan semua pihak yang terlibat dalam menyelesaikan skripsi ini.

Bandar Lampung, Juli 2018

Penulis

Annisa Rizki utami

DAFTAR ISI

DAFTAR TABEL

Tabel Halaman

DAFTAR GAMBAR

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Analisis regresi merupakan suatu metode dalam statistika yang mempelajari dan memodelkan pola hubungan antara variabel terikat (Y) dengan variabel bebas (X). Dalam regresi linier terdapat beberapa metode estimasi parameter, salah satunya adalah metode kuadrat terkecil. Pendugaan parameter pada metode ini diperoleh dengan meminimumkan jumlah kuadrat galat. Namun, metode ini sangat rentan terhadap adanya pencilan.

Selain itu, terdapat beberapa asumsi yang harus dipenuhi untuk melakukan estimasi parameter dengan metode kuadrat terkecil diantaranya adalah data harus mengikuti sebaran normal, homoskedastisitas, tidak ada multikolinearitas dan tidak ada autokorelasi. Pendugaan parameter dengan metode kuadrat terkecil yang memenuhi syarat asumsi klasik memiliki sifat *Best Linear Unbiased Estimator (BLUE)*.

Pada analisis regresi dengan kasus adanya pencilan yang menyebabkan asumsi klasik tidak terpenuhi, maka pendugaan parameter dengan metode kuadrat terkecil tidak dapat digunakan karena tidak bersifat *BLUE*, sehingga dibutuhkan metode alternatif untuk menduga parameter tanpa terpengaruh adanya pencilan.

Regresi kuantil merupakan suatu pendekatan dalam analisis regresi yang dikenalkan oleh Koenker dan Bassett (1978). Regresi kuantil dapat digunakan untuk mengatasi keterbatasan regresi linear dalam menganalisis asumsi yang tidak terpenuhi pada regresi klasik, yaitu galat tidak berdistribusi normal, mudah terpengaruh oleh data pencilan dan varians galat tidak konstan.

Oleh karena itu, berdasarkan hal-hal tersebut akan diilakukan analisis terhadap data simulasi yang telah terdeteksi adanya pencilan dengan menggunakan regresi kuantil.

1.2 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Menerapkan metode Regresi Kuantil pada data simulasi yang terdeteksi masalah pencilan.
2. Mengetahui keakuratan metode Regresi Kuantil dalam memprediksi model.
3. Membandingkan hasil prediksi model MKT dengan Regresi Kuantil.

1.3 Manfaat Penelitian

Manfaat dari penelitian ini adalah menambah pengetahuan bagi penulis dan memberi masukan kepada para peneliti dan pembaca tentang metode Regresi Kuantil untuk menganalisis data yang terdeteksi masalah pencilan.

II. TINJAUAN PUSTAKA

2.1 Analisis Regresi

Analisis regresi merupakan suatu teknik dalam statistika untuk memodelkan dan menyelidiki hubungan antara sebuah variable terikat (Y) dengan salah satu atau lebih variabel bebas ($X_1, X_2, X_3, \ldots, X_k$). Dalam regresi terdapat dua jenis model regresi yang terkenal, yaitu model regresi linier sederhana yang mengkaji dua variabel yaitu satu variabel terikat (Y) dan satu variabel bebas (X) dan model regresi linier berganda yang melibatkan dua atau lebih variabel bebas atau sering disebut juga dengan regresi klasik (Gujarati, 2003).

Jika terdapat vektor input **xT** = (**x1**, **x2**,…,**x**) dan digunakan untuk menduga luaran nilai Y yangberupa bilangan riil, maka model regresi linier memiliki bentuk sebagai berikut,

$$Y_i = \beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j + \varepsilon_i \qquad (2.1)$$

dengan

Y_i : Vektor variabel terikat berukuran nx1

β_0 : Intersep

x_{ij} : Matriks variabel bebas berukuran nx(p+1)

β_j : Slope atau kemiringan

ε_i : Vektor galat berukuran nx1

2.2 Model Regresi Linier Berganda

Model regresi linier berganda merupakan perluasan dari model regresi linier sederhana. Dengan memperluas hubungan secara linier antara dua atau lebih variabel bebas dengan satu variabel terikat. Misalkan n > k observasi, dan x_{ij} dari pengamatan ke- i observasi. Model pengamatannya adalah

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + \varepsilon_i \tag{2.2}$$

dengan

i : 1, 2, 3, ..., n

k : 1, 2, 3, ..., n

Y_i : Variabel terikat pengamatan ke-i

X_{ki} : Variabel bebas pengamatan ke-i

β_0 : Konstanta (parameter)

β_k : Koefisien regresi atau *slope*(parameter) ke-k

ε_i : Sisaan (galat) pengamatan ke-i

Dari persamaan tersebut dapat dituliskan dalam notasi matriks dengan persamaan sebagai berikut:

$$\boldsymbol{Y}=\boldsymbol{X\beta}+\boldsymbol{\varepsilon} \text{ dengan } \boldsymbol{\varepsilon} \sim N(0, \sigma^2{}_I) \tag{2.3}$$

dengan

Y : Vektor n x 1 variabel terikat

X : Matriks n x k variabel bebas

β : Vektor k x 1 koefisien variabel bebas

£ : Vektor n x 1 variabel acak galat dengan E(ε) = 0 dan matriks ragam peragam $\sigma^2(\varepsilon) = \sigma^2 I$

2.3 Metode Kuadrat Terkecil (MKT)

Estimasi parameter bertujuan untuk mendapatkan model regresi yang akan digunakan dalam analisis regresi. Salah satu metode yang dapat digunakan untuk mengestimasi parameter model regresi linier sederhana maupun model regresi linier berganda adalah dengan metode kuadrat terkecil (MKT).

Metode kuadrat terkecil merupakan metode yang digunakan untuk menduga koefisien regresi linear dengan cara meminimumkan jumlah kuadrat galat (Hastie, 2008). Menurut Sembiring (1995), untuk mencari nilai-nilai β yaitu dengan meminimumkan bentuk kuadrat

$$Q(\beta_j) = (Y - X\beta)'(Y - X\beta)$$

$$= Y'Y - 2Y'X\beta + \beta'X'X\beta \qquad (2.4)$$

kemudian dicari turunan dari $Q(\beta_j)$ secara parsial terhadap $\beta_j;\ j = 1,2,\dots,k$ dan disamakan dengan nol.

$$\frac{\partial Q}{\partial \beta o} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_{i1} - \beta x_{i2} - \dots - \beta_k x_{ik}) = 0$$

$$\frac{\partial Q}{\partial \beta 1} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \dots - \beta_k x_{ik})x_{i1} = 0$$

$$\frac{\partial Q}{\partial \beta 2} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \dots - \beta_k x_{ik})x_{i2} = 0$$

$$\frac{\partial Q}{\partial \beta k} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \dots - \beta_k x_{ik})x_{ik} = 0$$

Setelah disusun kembali dan mengganti semua parameter dengan penduganya, maka sistem persamaan diatas dapat ditulis sebagai

$$\sum_{i=1}^{n} y_i = nb_0 + b_1\sum_{i=1}^{n} x_{i1} + b_2\sum_{i=1}^{n} x_{i2} + \dots + b_k\sum_{i=1}^{n} x_{ik}$$
$$\sum_{i=1}^{n} x_{i1}\, y_i = b_0\sum_{i=1}^{n} x_{i1} + b_1\sum_{i=1}^{n} x_{i2}^2 + b_2\sum_{i=1}^{n} x_{i2}x_{i1} + \dots + b_k\sum_{i=1}^{n} x_{ik}\, x_{i1}$$
$$\sum_{i=1}^{n} x_{i2}\, y_i = b_0\sum_{i=1}^{n} x_{i2} + b_1\sum_{i=1}^{n} x_{i2}x_{i1} + b_2\sum_{i=1}^{n} x_{i2}^2 + \dots + b_k\sum_{i=1}^{n} x_{ik}\, x_{i2}$$
$$\sum_{i=1}^{n} x_{ik}\, x_{ik} = b_0\sum_{i=1}^{n} x_{ik} + b_1\sum_{i=1}^{n} x_{i1}\, x_{ik} + b_2\sum_{i=1}^{n} x_{i2}\, x_{ik} + \dots + b_k\sum_{i=1}^{n} x^2{}_{ik}$$

Persamaan ini merupakan persamaan normal. Jika ditulis dalam bentuk matriks maka bentuknya menjadi

$$\begin{bmatrix} n & \sum_{i=1}^{n} X_{i1} & \sum_{i=1}^{n} X_{i2} & \dots & \sum_{i=1}^{n} X_{ik} \\ \sum_{i=1}^{n} X_{i1} & \sum_{i=1}^{n} X_{i1}^2 & \sum_{i=1}^{n} X_{i2}X_{i1} & \dots & \sum_{i=1}^{n} X_{ik}X_{i1} \\ \sum_{i=1}^{n} X_{i2} & \sum_{i=1}^{n} X_{i1}X_{i2} & \sum_{i=1}^{n} X_{i2}^2 & \dots & \sum_{i=1}^{n} X_{ik}X_{i2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} X_{ik} & \sum_{i=1}^{n} X_{i1}X_{ik} & \sum_{i=1}^{n} X_{i2}X_{ik} & \dots & \sum_{i=1}^{n} X_{ik}^2 \end{bmatrix} \begin{bmatrix} b_o \\ b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_{11} & X_{21} & \dots & X_{n1} \\ X_{12} & X_{22} & \dots & X_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{ik} & X_{2k} & \dots & X_{nk} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix}$$

Atau secara lengkap jika ditulis kedalam bentuk matriks menjadi

$$X^T X \hat{\beta} = X^T Y \qquad (2.5)$$

Bila X'X tidak singular maka ada inversnya, sehingga diperoleh penduga untuk MKT

$$\hat{\beta} = (X^T X)^{-1} X^T Y \qquad (2.6)$$

Sifat-sifat penduga metode kuadrat terkecil adalah sebagai berikut

1. $\hat{\beta}$ linier

   $\hat{\beta}$ linier jika $\hat{\beta}$ merupakan fungsi linier dari β

$$\begin{aligned} \hat{\beta} &= (X^T X)^{-1} X^T Y \\ &= (X^T X)^{-1} X^T (X\beta + \varepsilon) \\ &= (X^T X)^{-1} X^T X\beta + (X^T X)^{-1} X^T \varepsilon \\ &= I\beta + (X^T X)^{-1} X^T \varepsilon \end{aligned}$$

2. $\hat{\beta}$ tak bias

$\hat{\beta}$ adalah penduga tak bias jika $E(\hat{\beta}) = \beta$

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

$$= E((X^T X)^{-1} X^T X\beta + \varepsilon)$$

$$= E((X^T X)^{-1} X^T X\beta + (X^T X)^{-1} X^T \varepsilon)$$

$$= (X^T X)^{-1} X^T X\beta + (X^T X)^{-1} X^T E(\varepsilon)$$

$$= (X^T X)^{-1} X^T X\beta$$

$$= \beta$$

Sehingga $\hat{\beta}$ merupakan penduga tak bias dari β

3. $\hat{\beta}$ memiliki variansi minimum

$$var(\hat{\beta}) = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T]$$

$$= E[((X^T X)^{-1} X^T Y - \beta)((X^T X)^{-1} X^T Y - \beta)^T]$$

$$= E[\ ((X^T X)^{-1} X^T (X\beta + \varepsilon) - \beta)((X^T X)^{-1} X^T (X\beta + \varepsilon) - \beta)^T]$$

$$= E[((X^T X)^{-1} X^T X\beta + (X^T X)^{-1} X^T \varepsilon) - \beta)((X^T X)^{-1} X^T X\beta + (X^T X)^{-1} X^T \varepsilon) - \beta)^T]$$

$$=E[(\beta + (X^T X)^{-1} X^T \varepsilon - \beta)(\beta + (X^T X)^{-1} X^T \varepsilon - \beta)^T]$$

$$=E[((X^T X)^{-1} X^T \varepsilon)((X^T X)^{-1} X^T \varepsilon)^T]$$

$$=E[((X^T X)^{-1} X^T \varepsilon)((X^T X)^{-1} X^T \varepsilon)^T]$$

$$=E[((X^T X)^{-1} X^T \varepsilon^T \varepsilon\ X\ (X^T X)^{-1}]$$

$$=E[((X^T X)^{-1} X^T\ E(\varepsilon\varepsilon^T)\ X\ (X^T X)^{-1}]$$

$$=(X^T X)^{-1} X^T X\ (X^T X)^{-1}\ E[\varepsilon\varepsilon^T]$$

$$=(X^T X)^{-1} \sigma^2$$

$Var(\hat{\beta}) = (X^T X)^{-1}\sigma^2$ merupakan varians terkecil dari semua penaksir linier tak bias.

Estimator kuadrat terkecil yang memenuhi sifat linear, tak bias, dan mempunyai variansi minimum ini bersifat *Best Linear Unbiased Estimator* (*BLUE).*

2.4 Asumsi Klasik

Asumsi klasik adalah persyaratan statistik yang harus dipenuhi pada analisis regresi linier berganda yang berbasis Metode Kuadrat Terkecil (MKT).Jadi analisis regresi yang tidak berdasarkan MKT tidak memerlukan persyaratan asumsi klasik, misalnya regresi logistik atau regresi ordinal. Demikian juga tidak semua uji asumsi klasik harus dilakukan pada analisis regresi linier, misalnya uji multikolinearitas tidak dilakukan pada analisis regresi linier sederhana dan uji autokorelasi tidak perlu diterapkan pada data *cross sectional*. Berikut adalah uji asumsi klasik pada regresi linier berganda.

1. Uji Linearitas

Uji linearitas dipergunakan untuk melihat apakah model yang dibangun mempunyai hubungan linier atau tidak.Uji ini jarang digunakan pada berbagai penelitian, karena biasanya model dibentuk berdasarkan telaah teoretis bahwa hubungan antara variabel bebas dengan variabel terikatnya adalah linier.

Uji linearitas digunakan untuk mengkonfirmasikan apakah sifat linier antara dua variabel yang diidentifikasikan secara teori sesuai atau tidak dengan hasil observasi yang ada. Uji linearitas dapat menggunakan uji Durbin-Watson, Ramsey Test atau uji Lagrange Multiplier.

2. Uji Normalitas

Uji normalitas adalah untuk melihat apakah nilai residual terdistribusi normal atau tidak. Model regresi yang baik adalah memiliki nilai residual yang terdistribusi normal. Jadi uji normalitas bukan dilakukan pada masing-masing variabel tetapi pada nilai residualnya. Sering terjadi kesalahan yang jamak yaitu bahwa uji normalitas dilakukan pada masing-masing variabel. Hal ini tidak dilarang tetapi model regresi memerlukan normalitas pada nilai residualnya bukan pada masing-masing variabel penelitian.

Uji normalitas dapat dilakukan dengan uji histogram, uji normal P Plot, uji Chi Square, Skewness dan Kurtosis atau uji Kolmogorov Smirnov. Tidak ada metode yang paling baik atau paling tepat. Tipsnya adalah bahwa pengujian dengan metode grafik sering menimbulkan perbedaan persepsi di antara beberapa pengamat, sehingga penggunaan uji normalitas dengan uji statistik bebas dari keragu-raguan, meskipun tidak ada jaminan bahwa pengujian dengan uji statistik lebih baik dari pada pengujian dengan metode grafik.

3. Uji Homoskedastisitas

Uji heteroskedastisitas adalah untuk melihat apakah terdapat ketidaksamaan varians dari residual satu pengamatan ke pengamatan yang lain. Model regresi yang memenuhi persyaratan adalah di mana terdapat kesamaan varians dari residual satu pengamatan ke pengamatan yang lain tetap atau disebut homoskedastisitas.

Deteksi heteroskedastisitas dapat dilakukan dengan metode scatter plot. Model yang baik didapatkan jika tidak terdapat pola tertentu pada grafik, seperti mengumpul di tengah, menyempit kemudian melebar atau sebaliknya. Uji statistik yang dapat digunakan adalah uji Glejser, uji Park atau uji White, dll.

4. Uji Autokorelasi

Uji autokorelasi adalah untuk melihat apakah terjadi korelasi antara suatu periode t dengan periode sebelumnya (t -1). Secara sederhana adalah bahwa analisis regresi adalah untuk melihat pengaruh antara variabel bebas terhadap variabel terikat, jadi tidak boleh ada korelasi antara observasi dengan data observasi sebelumnya.

Uji autokorelasi hanya dilakukan pada data runtun waktu dan tidak perlu dilakukan pada data *cross section* seperti pada kuesioner di mana pengukuran semua variabel dilakukan secara serempak pada saat yang bersamaan.

Beberapa uji statistik yang sering dipergunakan adalah uji Durbin-Watson, uji dengan Run Test dan jika data observasi di atas 100 data sebaiknya menggunakan uji Lagrange Multiplier.

Uji Durbin-Watson akan menghasilkan nilai Durbin-Watson (dw) yang nantinya akan dibandingkan dengan dua nilai Durbin-Watson tabel, yaitu Durbin Upper (dU) dan Durbin Lower (dL). Cara menentukan atau kriteria pengujian autokorelasi berdasarkan nilai Durbin-Watsonadalah sebagai berikut.

Jika dw< dL maka terdapat autokorelasi positif,

Jika dw> dU maka tidak terdapat autokorelasi positif,

Jika dL <dw< dU maka pengujian tidak meyakinkan atau tidak dapat disimpulkan.

Jika (4 – dw) < dL maka terdapat autokorelasi negatif,

Jika (4 – dw) > dU maka tidak terdapat autokorelasi negatif,

Jika dL < (4 – dw) < dU maka pengujian tidak meyakinkan atau tidak dapat disimpulkan.

Dikatakan tidak terdapat autokorelasi jika nilai dw> dU dan (4-dU) > dw atau bisa dinotasikan juga sebagai berikut: (4-dU) > dw > dU.

5. Uji Multikolinearitas

Uji multikolinearitas adalah untuk melihat ada atau tidaknya korelasi yang tinggi antara variabel-variabel bebas dalam suatu model regresi linear berganda. Jika ada korelasi yang tinggi di antara variabel-variabel bebasnya, maka hubungan antara variabel bebas terhadap variabel terikatnya menjadi terganggu.

Alat statistik yang sering dipergunakan untuk menguji gangguan multikolinearitas adalah dengan *variance inflation factor* (VIF), korelasi pearson antara variabel-variabel bebas, atau dengan melihat nilai eigen dan *condition index* (CI).

2.5 Pencilan

Menurut Barnett (1981), pencilan adalah pengamatan yang tidak mengikuti sebagian besar pola danterletak jauh dari pusat data. Menurut Ferguson (1961), pencilan didefinisikan sebagai suatu data yang menyimpang dari sekumpulan data yang lain. Menurut R.K Sembiring (1950) Pencilan adalah pengamatan yang jauh dari pusat data yang mungkin berpengaruh besar terhadap koefesien regresi (Soemartini, 2007).

2.6 Kuantil

Kuantil adalah nilai-nilai observasi yang membagi data menjadi N bagian yang sama. Menghitung nilai-nilai kuantil suatu data yang disajikan dalam distribusi frekuensi, sama dengan cara menghitung median, dengan rumus :

$$\mathrm{Ki} = \mathrm{Li} + \frac{\left(\frac{\mathrm{i}}{\mathrm{r}}\right)\mathrm{n} - \mathrm{F}}{\mathrm{f}} \text{ x c}$$

dengan

Ki : Kuantil ke-i

Li : Batas bawah kelas kuantil ke-i

n : Banyaknya data

F : Banyaknya frekuensi seluruh kelas yang lebih rendah dari kelas kuantil ke-i

f : Frekuensi kelas kuantil ke -i

c : Lebar interval kelas kuantil

2.7 Regresi Kuantil

Regresi kuantil merupakan suatu pendekatan dalam analisis regresi yang dikenalkan oleh Koenker dan Bassett (1978). Pendekatan ini menduga berbagai fungsi kuantil dari suatu distribusi Y sebagai fungsi dari X. Regresi kuantil sangat berguna jika distribusi data tidak homogen dan tidak berbentuk standar seperti tidak simetris, terdapat ekor pada sebaran, atau *truncated distribution*. Misalkan Y adalah peubah acak dengan fungsi distribusi F_Y dan τ adalah konstanta dimana $0 < \tau < 1$. Kuantil ke- τ dari F_Y dinotasikan sebagai $q_Y(\tau)$ solusi untuk $F_Y(q) = \tau$, yaitu:

$$q_Y(\tau) = F_Y^{-1}(\tau) = inf\{y{:}F_Y(y) \geq \tau\}$$

Sehingga $100\tau\%$ $(100\,(1-\tau)\%)$ dari masa peluang Y berada di bawah (di atas) $q_Y(\tau)$.

Seperti halnya dengan metode MKT yang meminimumkan jumlah kuadrat sisaan untuk mencari nilai dugaan bagi β, maka dalam regresi kuantil, kuantil ke-τ dari F_Y dapat diperoleh dengan meminimumkan fungsi berikut ini terhadap q:

$$\tau \int_{y>q} |y-q| dF_Y(y) + (1-\tau)\int_{y<q} |y-q| dF_Y(y) =$$

$$\tau \int_{y>q} |y-q| dF_Y(y) - (1-\tau)\int_{y<q} |y-q| dF_Y(y) \qquad (2.7)$$

Dengan meminimumkan fungsi di atas, akan diperoleh persamaan berikut:

$$0 = -\tau \int_{y>q} dF_Y(y) + (1-\tau)\int_{y>q} dF_Y(y)$$

$$= -\tau\,[1 - F_Y(q)] + (1-\tau)F_Y(q)$$

$$= -\tau + F_Y(q)$$

Sehingga kuantil ke τ merupakan solusi dari F_Y.

$$\tau = F_Y(q) \tag{2.8}$$

Jika Y sebagai fungsi dari X yang telah diketahui, memiliki peluang $F_{y|x}(y)$, kuantil ke-τ dari fungsi tersebut dapat dituliskan sebagai $Q_{Y|X}(\tau) = F_{Y|X}^{-1}(\tau)$. $Q_{Y|X}(\tau)$ merupakan fungsi dari X dan diselesaikan dengan persamaan berikut:

$$\text{Min}_q \tau \int_{y>q} |y - q| dF_Y(y) + (1 - \tau) \int_{y<q} |y - q| dF_Y(y) \tag{2.9}$$

$Q_{Y|X}(0.5)$ adalah median Y (sebagai fungsi dari X) yang menunjukkan titik simetri dari $F_{Y|X}$; untuk τ mendekati 0 (atau 1), $Q_{Y|X}(\tau)$ menunjukkan ekor kiri (atau kanan) dari $F_{Y|X}$. Dalam notasi matriks, jika $Q_{Y|X}(\tau)$ adalah fungsi linear **X'β**, maka persamaan (2.9) akan menjadi:

$$\text{Min}_q \tau \int_{y>X'\beta} |y - X'\beta| dF_Y(y) + (1 - \tau) \int_{y<q} |y - X'\beta| dF_Y(y) \tag{2.10}$$

Solusi dari persamaan 2 ini dinotasikan sebagai βτ dan kuantil Y (sebagai fungsi dari X) ke-τ adalah $Q_{Y|X}(\tau) = X'\beta_\tau$

Misalnya diberikan data **(y_t, x_t)** untuk t=1, 2, …, T, dimana $\mathbf{x_t}$ berukuran **k x 1**, maka model linier dari persamaan regresi kuantil dapat dituliskan sebagai:

$$y_t = x_t'\beta + \varepsilon_t \tag{2.11}$$

dengan $Q_\tau(y_t|x_t) = x_t'\beta$ merupakan kuantil ke-τ (0 < τ < 1) dari y dengan suatu nilai tertentu. Penduga bagi β dari regresi kuantil ke-τ diperoleh dengan meminimumkan jumlah nilai mutlak dari galat dengan pembobot τ untuk galat positif dan pembobot (1- τ) untuk galat negatif yaitu:

$$\hat{\beta} = min_\beta\{\tau \sum_{t|y_t \geq x_t} |y_t - x_t'\beta| + (1 - \tau) \sum_{t|y_t < x_t} |y_t - x_t'\beta| \} \tag{2.12}$$

atau

$$\hat{\beta} = min_{\beta} \frac{1}{T} \sum_{i=1}^{T} \rho_{\tau}\, u_t \qquad (2.13)$$

dengan $\rho_{\tau}(u_t) = \begin{cases} \tau u_t & ,jika\ u_t \geq 0 \\ (\tau - 1)u_t & ,jika\ u_t < 0 \end{cases}$

$\rho_{\tau} u_t$ disebut juga sebagai *Check Function* dan galat dugaan dari y adalah $\hat{\varepsilon} = y_t - x_t'\beta$. Solusi dari persamaan (2.13) tidak dapat diperoleh secara analitik, tetapi dilakukan secara numerik. Salah satu metode yang dapat digunakan untuk menyelesaikan persamaan (2.13) adalah metode *simpleks*.

2.8 *Mean Square Error* (MSE)

Jika $\hat{\beta}$ adalah penduga yang tak bias dari β, maka E $((\hat{\beta}) - \beta)^2$ sama dengan ragam penduga $\hat{\beta}$. Tetapi jika suatu $\hat{\beta}$ adalah penduga yang bias dari β, maka E $((\hat{\beta}) - \beta)^2$ disebut dengan MSE dari $\hat{\beta}$.

MSE $(\hat{\beta})$ = E $((\hat{\beta}) - \ )^2$

Bukti:

$$\begin{aligned} \text{MSE}(\hat{\beta}) &= \text{E}(\hat{\beta}^2 - 2\hat{\beta}\,\beta + \beta^2) \\ &= \text{E}(\hat{\beta}^2) - 2\text{E}(\hat{\beta})\,\beta + \beta^2 \\ &= \{\text{E}(\hat{\beta}^2) - 2\text{E}(\hat{\beta})\,\beta + \beta^2\} + \{(\text{E}(\hat{\beta}))^2 - (\text{E}(\hat{\beta}))^2\} \\ &= \{\text{E}(\hat{\beta}^2) - (\text{E}(\hat{\beta}))^2\} + \{(\text{E}(\hat{\beta}))^2 - 2\text{E}(\hat{\beta})\,\beta + \beta^2\} \\ &= \{\text{E}(\hat{\beta}^2) - (\text{E}(\hat{\beta}))^2\} + \{(\text{E}(\hat{\beta}) - \beta)^2\} \\ &= \text{var}(\hat{\beta}) + (\text{bias}(\hat{\beta}))^2 \end{aligned}$$

(Suhupi, 2006).

2.9 *Akaike Information Criterion* (AIC)

AIC adalah standar informasi yang menyediakan ukuran informasi yang dapat menemukan keseimbangan antara ukuran kebaikan model dan spesifikasi model. Model yang baik adalah model yang memiliki nilai AIC yang terkecil, dengan rumus AIC adalah sebagai berikut

$$AIC = \ln(MSE) + 2*K/N$$

dengan

MSE : Mean Squared Error

K : Jumlah parameter yang diestimasi

N : Jumlah observasi

(Junaidi, 2013).

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilakukan pada semester ganjil tahun ajaran 2017/2018. Bertempat di jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.

3.2 Data Penelitian

Data yang digunakan dalam penelitian ini adalah data simulasi dari model regresi linear berganda dengan variabel bebas (p) sebanyak 3 dimana $n =$ 30 dan 50 dengan model sebagai berikut:

$$Y_i = X_{i1} + X_{i2} + X_{i3} + \varepsilon_i$$

untuk $i = 1, 2, \dots, n$. Koefisien regresi ditetapkan yaitu $\beta_0 = 0$ dan $\beta_1 = \beta_2 = \beta_3 = 1$. Seluruh variabel bebas dibangkitkan sebanyak n dari distribusi N(0,1) dengan kontaminasi pencilan sebanyak 10%, 20%, 30%, 40%, dan 50% dari distribusi N(15,1).

3.3 Metode Penelitian

Penelitian ini dilakukan dengan menggunakan studi literatur secara sistematis yang diperoleh dari buku-buku maupun media lain untuk mendapatkan informasi sebanyak mungkin untuk mendukung penulisan proposal penelitian ini. Analisis data dilakukan dengan software R versi 3.4.2. Adapun langkah-langkah yang dilakukan adalah:

1. Membangkitkan data untuk masing-masing variabel bebas X_1, X_2 dan X_3 dengan $n =$ 30 dan 50 dari sebaran N(0,1) dengan kontaminasi pencilan sebanyak 10%, 20%, 30%, 40%, dan 50% dari distribusi N(15,1) .
2. Menetapkan seluruh koefisien regresi, yaitu $\beta_0 = 0$ dan $\beta_1 = \beta_2 = \beta_3 = 1$.
3. Menentukan nilai y dari model berikut:

$$Y_i = X_{i1} + X_{i2} + X_{i3} + \varepsilon_i$$

4. Melakukan estimasi parameter dengan metode MKT.
5. Melakukan uji asumsi normalitas yang bertujuan untuk mengetahui apakah sifat *BLUE* pada metode ini terpenuhi atau tidak.
6. Melakukan estimasi parameter dengan metode regresi kuantil, yaitu dengan algoritma simpleks.

   Sebagai langkah awal misalkan, μ= $[y - A'\beta]_+$, v= $[A'\beta - y]_+$ +, ϕ= $[\beta]_+$, dan φ= $[-\beta]_+$, dimana $[z]_+$ adalah bagian dari z yang tidak bernilai negatif, dan A merupakan matriks peubah penjelas. Untuk kasus regresi median, pendekatan simpleks menyelesaikan $min_\beta D_{LAR}(\beta)$dengan memformulasikan

$$min_\beta \{e'\mu{+}e'v|y{=}A'\beta{+}\mu{-}v\}$$

   di mana e merupakan vektor satu yang berukuran n dan $\{u,v\} \in \mathbb{R}^n_+$. Misalkan B= $[A'{-}A'I{-}I]$, d=$(0'\ 0'\ e'\ e')$, θ=$(\phi'\ \varphi'\ \mu'\ v')$ di mana0′=(0 0…0)p. Sehingga rumusan ulang dari masalah pemrograman linier baku adalah min_θ d'θ dengan kendala Bθ = y dan θ ≥ 0. Masalah ini memiliki bentuk ganda max_Z x'z dengan kendala B'z = d. Yang dapat disederhanakan menjadi:

$$max_Z\{y'z|Az = 0, z \in [-1,1]^n\}$$

Jika $\eta = \frac{1}{2} z + \frac{1}{2} e$, $b = \frac{1}{2} Ae$, maka rumusannya menjadi:

$$max_\eta\{y'\eta|A\eta = b, \eta \in [0,1]^n\}$$

Untuk regresi kuantil, masalah minimisasinya adalah $min_{\beta \in \mathbb{R}^P} \sum_{i=1}^{n} \rho_\tau(y_i - x_i^T\beta)$, dan sama seperti tahapan sebelumnya, rumusan masalahnya menjadi

$$max_Z\{y'z|Az = (1\text{-}\tau), z \in [0,1]^n\}$$

7. Membandingkan nilai duga β dan model dengan menggunakan MSE dan nilai *AIC*.
8. Membuat kesimpulan dari metode terbaik.

V. KESIMPULAN

Dari hasil-hasil yang telah didapat, kesimpulan tentang analisis regresi dengan metode Regresi Kuantil dapat dikemukakan sebagai berikut.

1. Metode Regresi Kuantil baik digunakan pada data dengan kontaminasi pencilan kurang dari sama dengan 50% dari jumlah data.
2. Berdasarkan uji kebaikan model, nilai MSE dan AIC yang diperoleh dari metode regresi kuantil lebih kecil dibandingkan dengan MKT, sehingga metode regresi kuantil lebih baik digunakan untuk menduga nilai beta daripada MKT pada data yang mengandung pencilan.
3. Kuantil terbaik bagi model regresi pada data yang mengandung pencilan berada pada kuantil 0.50, karena nilai MSE dan AIC paling kecil berada pada kuantil 0.50.

DAFTAR PUSTAKA

Barnett,V. and Lewis,T. 1984. *Outlier in Statistical Data*. John Wiley & Sons, New York.

Junaidi, E. 2013. *Analisis volatilitas harga minyak sawit dan harga minyak goreng.* Sekolah Pascasarjana IPB, Bogor.

Furno, M. 2007. *Parameter Instability in Quantile Regressions*. Elsevier, New York.

Gujarati, N. D. 2003. *Basic Econometrics*. Ed. ke-4. McGraw-Hill Companies Incorporation, New York.

Hastie T, Tibshirani R, and Friedman J. 2008. *The Elements of Statistica Learning. Data Mining, Inference, and Prediction*. Ed. ke-2. Springer, New York.

Koenker, R. and Basset. 1978. *Regression Quantile Econometrica*. Cambridge University Press, London.

Koenker, R. and Hallock, K.F. 2001. *Quantile Regression.* American Economic Association, New York.

Kuan, C.M. 2007. *An Intriduction to Quantile Regression*. Institute of Economics, Academia Sinica. Canada.

Kusrini, S.D.E. 2010. *Ekonometrika.* Andi Yogyakarta, Yogyakarta.

Kutner, M. H., C.J. Nachtsheim., and J. Neter. 2004. *Applied Linear Regression Models*. Ed. ke-4. McGraw-Hill Companies Incorporation, New York.

Rawlings, J. O., Pantula, and S. G., Dickey, D. A. 1998. Applied *Regression Analysis: a Research Tool*. Ed ke-2. Springer-Verlag, Inc ., New York.

Sembiring, R. K. 1995. *Analisis Regresi*. Ed Ke-2. ITB, Bandung.

Soemartini. 2007, Pencilan . http://www.math.itb.ac.id/ma291/pencilan.html. Diakses tanggal 05 Februari 2018.

Suhupi, A. 2006. Pendugaan Matriks Kovarian Robust Dengan Metode MVE Untuk Mendeteksi Pencilan Pada Analisis Multivariat. Skripsi. Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung, Bandar Lampung.

Uthami, I. A. P., Sukarsa, I. K. G., dan Kencana, I. P. K. 2013. Regresi Kuantil Median untuk Mengatasi Heteroskedastisitas pada Analisis Regresi. *E-Jurnal Matematika* **2**(1), hal. 7-12.

Widarjono, A. 2007. *Ekonometrika*: *Teori dan Aplikasi untuk Ekonomi dan Bisnis*. Ed. ke-2. Jurusan Ekonisia Fakultas Ekonomi Universitas Islam Indonesia, Yogyakarta.