ESTIMASI MODEL REGRESI SEMIPARAMETRIK DENGAN PENDUGA NADARAYA-WATSON KERNEL UNIFORM

(Skripsi)

Oleh

RISWANTI OKTAVIANI

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2019

ABSTRACT

SEMIPARAMETRIC REGRESSION MODEL ESTIMATION WITH NADARAYA-WATSON UNIFORM KERNEL ESTIMATOR

By

RISWANTI OKTAVIANI

Semiparametric regression model is an estimation that combines both parametric and nonparametric regression model. This study aims to analyze semiparametric regression model using uniform kernel estimator with dengue fever case and simulated data. The optimal bandwidth of semiparametric regression curve is obtained based on GCV value. The result shows that semiparametric regression model with uniform kernel is able to estimate the real and simulated data very well.

Keywords*: Semiparametric Regression, Kernel, Bandwidth, GCV*

ABSTRAK

ESTIMASI MODEL REGRESI SEMIPARAMETRIK DENGAN PENDUGA NADARAYA-WATSON KERNEL UNIFORM

Oleh

RISWANTI OKTAVIANI

Estimasi model regresi semiparametrik adalah estimasi yang menggabungkan model regresi parametrik dan nonparametrik. Tujuan dari penelitian ini adalah untuk menganalisis model regresi semiparametrik dengan menggunakan estimator kernel uniform dari kasus demam berdarah dan data simulasi. Bandwidth optimal pada kurva regresi semiparametrik didasarkan pada kriteria GCV. Hasil dari penelitian ini menunjukkan bahwa model regresi semiparametrik dengan kernel uniform dapat mengestimasi data asli dan data simulasi dengan sangat baik.

Kata Kunci: Regresi Semiparametrik, Kernel, *Bandwidth*, GCV

ESTIMASI MODEL REGRESI SEMIPARAMETRIK DENGAN PENDUGA NADARAYA-WATSON KERNEL UNIFORM

Oleh

Riswanti Oktaviani

Skripsi
Sebagai salah satu syarat untuk mencapai gelar
SARJANA SAINS

Pada
Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam

JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2019

Judul Skripsi : **ESTIMASI MODEL REGRESI SEMIPARAMETRIK DENGAN PENDUGA NADARAYA-WATSON KERNEL UNIFORM**

Nama Mahasiswa : Riswanti Oktaviani

Nomor Pokok Mahasiswa : 1517031008

Jurusan : Matematika

Fakultas : Matematika dan Ilmu Pengetahuan Alam

MENYETUJUI

1. Komisi Pembimbing

Ir. Netti Herawati, M.Sc., Ph.D.	**Drs. Tiryono Ruby, M.Sc., Ph.D.**
NIP. 19650125 199003 2 001	**NIP. 19620704 198803 1 002**

2. Mengetahui

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 : **Drs. Tiryono Ruby, M.Sc., Ph.D.** .....

Penguji
Bukan Pembimbing : **Prof. Drs. Mustofa Usman, MA, Ph.D.** .....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Drs. Suratman, M.Sc.
NIP. 19640604 199003 1 002

Tanggal Lulus Ujian Skripsi : **15 Juli 2019**

PERNYATAAN SKRIPSI MAHASISWA

Saya yang bertanda tangan dibawah ini:

Nama : Riswanti Oktaviani

Nomor Pokok Mahasiswa : 1517031008

Judul : Estimasi Model Regresi Semiparametrik dengan Penduga Nadaraya-Watson Kernel Uniform

Jurusan : Matematika

Dengan ini menyatakan bahwa skripsi ini adalah hasil pekerjaan saya sendiri, bukan hasil orang lain, dan semua hasil tulisan yang tertuang dalam skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung. 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, 12 Juli 2019

Penulis,

Riswanti Oktaviani
NPM. 1517031008

RIWAYAT HIDUP

Penulis bernama lengkap Riswanti Oktaviani lahir di Bandar Lampung, pada tanggal 22 Oktober 1997. Penulis merupakan anak kedua dari dua bersaudara, pasangan Bapak Ari Widianto dan Ibu Sukini.

Penulis memulai pendidikan dari sekolah dasar di SD Negeri 2 Gedong Air dari tahun 2003. Kemudian melanjutkan ke pendidikan menengah pertama di SMP Negeri 10 Bandar Lampung pada tahun 2009. Pendidikan menengah atas di SMA Negeri 3 Bandar Lampung pada tahun 2012.

Penulis melanjutkan pendidikan di perguruan tinggi dan terdaftar sebagai mahasiswa di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Universitas Lampung (UNILA) melalui jalur Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN) tahun 2015. Sebagai bentuk penerapan ilmu perkuliahan, pada tahun 2018 penulis melakukan Kerja Praktik (KP) di Kantor Pelayanan Pajak Pratama Tanjung Karang Kantor Wilayah Direktorat Jenderal Pajak Bengkulu dan Lampung serta melaksanakan Kuliah Kerja Nyata (KKN) di Desa Way Kepayang, Kecamatan Kedondong, Kabupaten Pesawaran.

KATA INSPIRASI

"Sesungguhnya bersama kesulitan ada kemudahan."

(QS. Al-Insyirah : 6)

"*Hai orang-orang yang beriman, jadikanlah sabar dan shalat sebagai penolongmu, sesungguhnya Allah bersama orang-orang yang sabar*."

(QS. Al-Baqarah : 153)

"Hanya karena kamu gagal satu kali, bukan berarti kamu akan gagal terus menerus. Kamu hanya gagal ketika kamu berhenti mencoba."

(Anonim)

PERSEMBAHAN

Karyaku yang sederhana ini kupersembahkan kepada:

Bapak dan Ibu

Terima kasih kepada Bapak dan Ibu yang selalu mendoakan kesuksesanku, memberi semangat, nasihat, dukungan serta kasih sayang yang tiada henti.

Kakakku Kuswanto

Terima kasih kepada kakak yang selalu memberikan semangat dan keceriaan dalam hidupku.

Sahabat-sahabatku Putri, Nuy, Intan, dan Windi

Terima kasih kepada para sahabatku yang selalu memberikan semangat, doa, dan motivasi, serta kenangan indah selama ini.

Almamater dan Negeriku

SANWACANA

Puji dan syukur tak henti-hentinya tercurahkan kepada Allah SWT berkat segala rahmat dan hidayah-Nya sehingga skripsi dengan judul “Estimasi Model Regresi Semiparametrik dengan Penduga Nadaraya-Watson Kernel Uniform” dapat terselesaikan dengan baik dan tepat pada waktunya.

Terselesaikannya skripsi ini tidak lepas dari dukungan, bimbingan, saran, serta doa dari berbagai pihak. Oleh karena itu, dalam kesempatan ini diucapkan terimakasih kepada:

1. Ibu Ir. Netti Herawati, M.Sc., Ph.D., selaku dosen pembimbing satu sekaligus dosen pembimbing akademik yang selalu membimbing penulis dalam menyusun skripsi.
2. Bapak Drs. Tiryono Ruby, M.Sc., Ph.D., selaku pembimbing dua yang telah memberikan saran serta pembelajaran yang sangat bermanfaat dalam menyelesaikan skripsi.
3. Bapak Prof. Drs. Mustofa Usman, MA, Ph.D., selaku pembahas dan penguji skripsi yang telah memberikan evaluasi, arahan, dan saran demi perbaikan skripsi.
4. Ibu Prof. Dra. Wamiliana, M.A., Ph.D., selaku ketua jurusan matematika.

5. Bapak Drs. Suratman, M.Sc. selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
6. Seluruh dosen Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
7. Kedua orang tua tercinta, bapak dan ibu yang tiada henti-hentinya mendoakan putra-putrinya.
8. Kakakku tersayang, Kuswanto yang selalu menyemangati serta memberikan berbagai keceriaan.
9. Sahabat – sahabatku tersayang Putri, Nuy, Windi, Intan, dan Tira yang selalu setia mendengarkan keluh-kesah penulis selama menempuh perkuliahan.
10. Teman-teman satu bimbingan Azam, Neli, Dita, Meilinda, Irma, Farkhana, Wahyu, El, Lelvi, Rahma, Resti terima kasih atas semangat dan saran selama penyelesaian skripsi.
11. Teman - teman di Jurusan Matematika angkatan 2015.
12. Semua pihak yang terlibat dalam penyelesaian skripsi ini yang tidak dapat disebutkan satu per satu.

Penulisan skripsi ini masih jauh dari sempurna, sehingga informasi tambahan, saran, dan kritik untuk pengembangan lebih lanjut sangat diharapkan. Semoga skripsi ini dapat bermanfaat bagi kita semua.

Bandar Lampung, 12 Juli 2019
Penulis,

Riswanti Oktaviani

DAFTAR ISI

Halaman

DAFTAR GAMBAR xv

DAFTAR TABEL xvi

I. PENDAHULUAN 1

1.1 Latar Belakang dan Masalah 1
1.2 Tujuan Penelitian 2
1.3 Manfaat Penelitian 2

II. TINJAUAN PUSTAKA 4

2.1 Analisis Regresi 4
2.2 Regresi Parametrik 5
a. Uji Normalitas Galat 6
b. Uji Multikolinearitas 7
c. Uji Heteroskedatisitas 8
2.3 Regresi Nonparametrik 9
2.4 Estimator Densitas Kernel 10
2.5 Kernel Uniform 12
2.6 Ukuran Kebaikan *Bandwidth* Optimal 12
2.7 Regresi Semiparametrik 13
2.8 Estimasi Model Semiparametrik Kernel 14
a. *Partially Linear Model* 14
b. Estimasi Parameter Variabel Parametrik 15
c. Estimasi Kurva Regresi Variabel Nonparametrik 16
2.9 Ukuran Kebaikan Model 17

III. METODOLOGI PENELITIAN 18

3.1 Waktu dan Tempat Penelitian 18
3.2 Data Penelitian 18
3.3 Metode Penelitian 19

IV. HASIL DAN PEMBAHASAN 21

4.1 Penentuan Variabel Parametrik dan Variabel Nonparametrik pada Data Asli 21
4.2 *Partially Linear Model* pada Data Asli 23
4.2.1 Estimasi Parameter Regresi Variabel Parametrik 24
a. Uji Normalitas Galat 25
b. Uji Multikolinearitas 25
c. Uji Heteroskedatisitas Galat 26
4.2.2 Estimasi Kurva Regresi Nonparametrik 27
4.3 Hasil Estimasi Regresi Semiparametrik pada Data Asli 28
4.4 Penentuan Variabel Parametrik dan Variabel Nonparametrik pada Data Bangkitan 29
4.5 *Partially Linear Model* pada Data Bangkitan 32
4.5.1 Estimasi Parameter Regresi Variabel Parametrik 32
a. Uji Normalitas Galat 33
b. Uji Multikolinearitas 33
c. Uji Heteroskedatisitas Galat 34
4.5.2 Estimasi Kurva Regresi Nonparametrik 35
4.6 Hasil Estimasi Regresi Semiparametrik pada Data Bangkitan 37

V. KESIMPULAN 38

DAFTAR PUSTAKA 39

LAMPIRAN

DAFTAR GAMBAR

Gambar Halaman

2.1 Kurva Kernel Uniform 12

4.1 Diagram Pencar Suhu dengan Lama Sembuh Pasien 21

4.2 Diagram Pencar Hemoglobin dengan Lama Sembuh Pasien..... 22

4.3 Diagram Pencar Hematokrit dengan Lama Sembuh Pasien 23

4.4(a) Kurva Regresi Nonparametrik pada Data Asli dengan h = 1 28

4.4(b) Kurva Regresi Nonparametrik pada Data Asli dengan h = 1.6 28

4.4(c) Kurva Regresi Nonparametrik pada Data Asli dengan h = 2 28

4.5 Diagram Pencar X_1 dengan Y 30

4.6 Diagram Pencar X_2 dengan Y 30

4.7 Diagram Pencar X_3 dengan Y 31

4.8(a) Kurva Regresi Nonparametrik pada Data Bangkitan dengan h = 1 36

4.8(b) Kurva Regresi Nonparametrik pada Data Bangkitan dengan h = 1.6 36

4.8(c) Kurva Regresi Nonparametrik pada Data Bangkitan dengan h = 2 36

DAFTAR TABEL

Tabel Halaman

4.1 Nilai Estimasi Parameter Beta pada Data Asli 24

4.2 Uji *Kolmogorov-Smirnov* pada Data Asli 25

4.3 Tingkat Korelasi antarvariabel Bebas pada Data Asli..... 26

4.4 Uji *Breusch Pagan* pada Data Asli 26

4.5 Nilai *Bandwidth*, GCV, dan MSE pada Data Asli 27

4.6 Nilai Estimasi Parameter Beta pada Data Bangkitan 32

4.7 Uji *Kolmogorov-Smirnov* pada Data Bangkitan 33

4.8 Tingkat Korelasi antarvariabel Bebas pada Data Bangkitan..... 34

4.9 Uji *Breusch Pagan* pada Data Bangkitan 35

4.10 Nilai *Bandwidth*, GCV, dan MSE pada Data Bangkitan 35

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Analisis regresi merupakan salah satu metode statistika yang banyak digunakan, terutama untuk menjelaskan hubungan antara variabel tak bebas dengan variabel bebas. Analisis regresi terdiri dari dua berdasarkan jumlah variabel independen, yaitu analisis regresi linear sederhana dan regresi linear berganda. Analisis regresi sederhana merupakan hubungan antara dua variabel yaitu variabel independen dan variabel dependen. Sedangkan analisis regresi berganda merupakan hubungan antara lebih dari dua variabel, yaitu sekurang-kurangnya dua variabel independen dengan satu variabel independen.

Model regresi umunya dibagi ke dalam tiga bentuk, yaitu model regresi parametrik, model regresi nonparametrik, dan model regresi semiparametrik. Model regresi semiparametrik merupakan model yang mengkombinasikan model regresi parametrik dan regresi nonparametrik. Pada regresi semiparametrik, sebagai variabel penjelasnya bersifat parametrik dan sebagian lain bersifat nonparametrik. Regresi semiparametrik digunakan apabila pada hubungan antara

sekumpulan variabel bebas dan variabel terikat ada yang diketahui dan ada pula yang tidak diketahui (Fitriani, *et al.*, 2015).

Dalam regresi semiparametrik dibutuhkan teknik *smoothing.* Salah satu teknik *smoothing* adalah kernel uniform yang dalam regresi semiparametrik digunakan untuk menduga fungsi regresi yang tidak diketahui bentuk kurva regresinya. Penduga kernel fleksibel dan secara matematik mudah diselesaikan. Pada penduga kernel yang terpenting adalah pemilihan parameter pemulus (*bandwidth)* yang optimal untuk mendapatkan kurva regresi yang optimal. Salah satu metode untuk mencari nilai *bandwidth* yang optimal adalah dengan menggunakan kriteria *Generalized Cross Validation* (GCV). Pada penelitian ini akan dibahas bagaimana mengestimasi model regresi semiparametrik menggunakan estimator kernel uniform.

1.2 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah untuk mengestimasi model regresi semiparametrik menggunakan estimator kernel uniform.

1.3 Manfaat Penelitian

Berdasarkan hasil penelitian ini, diharapkan:

1. Dapat mengestimasi model regresi semiparametrik menggunakan estimator kernel uniform.

2. Meningkatkan pengembangan dan pengetahuan, khususnya mengenai regresi semiparametrik menggunakan estimator kernel uniform.
3. Dapat digunakan oleh pihak-pihak yang berkepentingan untuk menggunakan konsep dan gagasan baru yang dihasilkan penelitian mengenai pemodelan regresi semiparametrik menggunakan estimator kernel uniform.

II. TINJAUAN PUSTAKA

2.1 Analisis Regresi

Salah satu bagian statistika adalah analisis regresi. Analisis regresi yang pertama kali diperkenalkan oleh Sir Francis Galton yang dipergunakan sebagai konsep statistik. Analisis regresi merupakan salah satu teknik statistika yang sering digunakan untuk mengetahui hubungan kasual antara variabel tak bebas (*dependent variable*) dengan satu atau lebih variabel bebas (*independent variable*). Untuk data yang satu variabel bebas dan satu variabel tak bebas, analisis regresi disebut juga dengan analisis regresi linear sederhana. Sedangkan untuk data yang memiliki lebih dari satu variabel bebas dan satu variabel tak bebas disebut dengan analisis regresi linear berganda. Jadi, secara umum analisis regresi adalah metode analisis data yang menggambarkan hubungan antara variabel bebas dengan satu atau beberapa variabel tak bebas (Hosmer and Lemeshow, 2000).

Misalkan X adalah variabel prediktor dan Y adalah variabel respon untuk n pengamatan berpasangan $\{(xi, yi)\}_{i=1}^{n}$, maka hubungan linear antara kedua variabel dapat diketahui dengan model umum regresi yaitu:

$$y_i = m(x_i) + \varepsilon_i \,,\ i = 1, 2, \dots, n \qquad (1)$$

dengan y_i adalah variabel tidak bebas pada pengamatan ke-i, ε_i merupakan galat/sisaan/kesalahan ke-i yang diasumsikan menyebar normal dengan mean sama dengan 0 dan ragam tetap σ^2, serta $m(x_i)$ merupakan fungsi regresi atau kurva regresi (Marlin, 2016).

2.2 Regresi Parametrik

Regresi parametrik merupakan metode statistika yang digunakan untuk mengetahui pola hubungan antara variabel prediktor dengan variabel respon dengan asumsi bahwa telah diketahui bentuk fungsi regresinya. Hubungan antara variabel respon dan variabel prediktor dalam model dapat terjadi dengan fungsi linear maupun nonlinier dalam parameter (Winarti dan Sony, 2010).

Secara umum model regresi parametrik dengan variabel prediktor ke-*k* adalah:

$$y_i = \sum_{k=1}^{n} x_{ik}\beta_k + \varepsilon_i \qquad (2)$$

dengan,

x_i : nilai data ke-*i* dari variabel prediktor ke-*k*

i : 1, 2, …, n

y_i : variabel respon dari data ke-*i*

$\beta_1, \dots, \beta_k$: parameter yang tidak diketahui

ε_i : galat ke-*i* yang diasumsikan menyebar $N \sim (0, \sigma^2)$

Menurut Hardle (1994), model yang baik adalah yang memenuhi asumsi sebagai berikut:

a. Uji Normalitas Galat

Tujuan uji normalitas data ini adalah untuk mengetahui apakah dalam model regresi variabel pengganggu atau residual memiliki distribusi normal (Drapper dan Smith, 1992). Uji normalitas dapat dilakukan dengan menggunakan Uji *Kolmogorov-Smirnov*. Hipotesis dalam uji Uji *Kolmogorov-Smirnov* dapat dilakukan sebagai berikut:

H_0 : galat menyebar normal

H_1 : galat tidak menyebar normal

Menurut Corder dan Foreman (2009), statistik uji Kolmogorov-Smirnov adalah sebagai berikut:

$$D = |S(x) - F_0(x)|$$

dengan,

D : selisih antara fungsi distribusi frekuensi kumulatif

$S(x)$: fungsi distribusi frekuensi kumulatif sampel

$F_0(x)$: fungsi distribusi frekuensi kumulatif teoritis pada masing-masing kelas

Penerapan pada uji Kolmogorov-Smirnov adalah dengan membandingkan nilai D dengan D tabel. Jika nilai D tabel lebih besar dari D atau jika nilai signifikansi di atas 0,05 maka dengan taraf nyata 5% tidak cukup bukti untuk menolak H_0.

b. Uji Multikolinearitas

Multikolinearitas adalah suatu keadaan di mana terdapat korelasi antar variabel prediktor ketika dalam model regresi menggunakan lebih dari satu variabel prediktor. Apabila terjadi multikolinearitas pada data akan menyebabkan matriks $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ memiliki determinan sama dengan nol. Ada beberapa cara untuk mengetahui ada tidaknya multikolinearitas, diantaranya nilai korelasi (korelasi antarvariabel bebas), *variance inflation factor* (VIF) dan *tolerance*.

Uji multikolinearitas bertujuan untuk menguji apakan ditemukan adanya korelasi antarvariabel bebas (independen) pada model. Asumsi multikolinearitas mengharuskan adanya korelasi yang sempurna atau besar diantara variabel-variabel independen. Analisis koefisien korelasi bertujuan untuk mempelajari apakah ada hubungan antara dua variabel. Koefisien antarvariabel independen haruslah lemah. Jika korelasi kuat, terjadilah problem multikolinearitas. Koefisien korelasi dirumuskan sebagai berikut:

$$r_{x_1x_2} = \frac{n\sum_{i=1}^{n} x_1x_2 - (\sum_{i=1}^{n} x_1)(\sum_{i=1}^{n} x_2)}{\sqrt{\{n\sum_{i=1}^{n} {x_1}^2 - (\sum_{i=1}^{n} x_1)^2\}\{n\sum_{i=1}^{n} {x_2}^2 - (\sum_{i=1}^{n} x_2)^2\}}} \quad i = 1,2,3,\dots,n \qquad (3)$$

dimana x_1 dan x_2 adalah variabel bebas (independen) pada model, sedangkan n adalah banyaknya sampel yang digunakan (Santoso, 2012).

Metode untuk menguji adanya multikolinearitas dapat dilihat pula pada *tolerance value* atau *variance inflation factor* (VIF). Kedua ukuran ini menunjukkan setiap variabel independen manakah yang dijelaskan oleh variabel independen lainnya.

Tolerance mengukur variabilitas variabel independen yang terpilih tidak dijelaskan oleh variabel independen lainnya. Nilai *tolerance* yang rendah sama dengan nilai VIF yang tinggi. Nilai VIF dapat diperoleh dengan rumus berikut:

$$VIF = \frac{1}{1-{R_j}^2} \tag{4}$$

$$tol = \frac{1}{VIF} \tag{5}$$

dengan R_j adalah koefisien determinasi. Batas *tolerance value* adalah 0.10 dan nilai VIF adalah 10. Jika VIF > 10 dan nilai *tolerance* < 0.10, maka terjadi multikolinearitas pada penelitian tersebut (Santoso, 2012).

c. Uji Heteroskedatisitas Galat

Uji heteroskedatisitas bertujuan untuk menguji apakah dalam sebuah model regresi terjadi ketidaksamaan varians atau residual dari satu pengamatan ke pengamatan yang lain. Menurut Anselin (1988), uji heteroskedatisitas dapat dilakukan menggunakan pengujian *Breusch Pagan*. Hipotesis yang digunakan adalah sebagai berikut:

H_0: ${\sigma_1}^2 = {\sigma_1}^2$ (Tidak terjadi heteroskedatisitas)

H_1: ${\sigma_1}^2 \neq {\sigma_1}^2$ (Terjadi heteroskedatisitas)

Menurut Kutner (2004), uji statistik *Breusch Pagan* adalah sebagai berikut:

$$BP = \frac{(0{,}5)SSR}{(JKG/n)^2}$$

dengan,

BP : nilai statistik uji *Breusch Pagan*

SSR : *Sum of Square Regression* atau jumlah kuadrat dari selisih antara nilai prediksi dengan nilai rata-rata nilai prediksi

JKG : jumlah kuadrat galat

n : banyaknya data

Uji *Breusch Pagan* mengikuti sebaran *chi*-square. Penerapan pada uji *Breusch Pagan* adalah jika $BP < \chi^2_{(\alpha,db)}$ atau signifikansi di atas 0.05 maka tidak cukup bukti untuk menolak H_0 artinya dengan taraf nyata 5% varians galat data bersifat sama atau homogen (tidak terjadiheteroskedatisitas).

2.3 Regresi Nonparametrik

Metode regresi nonparametrik adalah metode analisis data yang sangat popular semenjak akhir tahun 1990-an. Kebutuhan terhadap analisis yang dapat menganalisis jenis data yang beragam dan belum tentu memenuhi asumsi-asumsi parametrik membuat teknik ini makin banyak digunakan. Karena sifatnya yang fleksibel, pendekatan model regresi nonparametrik berperan penting dalam memeriksa data longitudinal (Wu and Zhang, 2006).

Regresi nonparametrik merupakan suatu metode statistika yang digunakan untuk mengetahui pola hubungan antara variabel prediktor dengan respons ketika tidak diperoleh informasi sebelumnya tentang bentuk fungsi regresinya atau tidak diketahui bentuk kurva regresinya. Fungsi dari model regresi nonparametrik

dapat berbentuk apa saja, baik linear atau nonlinear. Misalkan variabel respon adalah y dan variabel predictor adalan *x* untuk *n* pengamatan, model umum regresi nonparametrik adalah:

$$y_i = m(x_i) + \varepsilon_i \tag{6}$$

dengan,

y_i : variabel respon dari data ke-*i*

$m(x_i)$: fungsi *smooth* ke-*i* yang tidak diketahui

ε_i : galat ke-*i* yang saling bebas diasumsikan menyebar $N \sim (0, \sigma^2)$

Metode regresi nonparametrik dapat digunakan dengan mengabaikan asumsi-asumsi yang melandasi metode regresi parametrik. Fungsi regresi nonparametrik diasumsikan mulus sehingga memiliki fleksibilitas yang tinggi dalam mengestimasi fungsi regresinya. Ada beberapa teknik dalam mengestimasi regresi nonparametrik yaitu salah satunya kernel.

2.4 Estimator Densitas Kernel

Estimator kernel merupakan pengembangan dari estimator histogram. Estimator diperkenalkan oleh Rosenblatt (1956) dan Parzen (1962) sehingga dapat disebut juga estimator densitas kernel Rosenblatt-Parzen (Eubank, 1998).

Secara umum kernel K dengan parameter pemulus (*bandwidth*) h didefinisikan sebagai:

$$K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right) \quad \text{untuk} - \infty < \mathrm{x} < \infty \text{ dan } \mathrm{h} > 0 \tag{7}$$

Serta memenuhi:

(i) K(x) $\geq$ 0, untuk semua x

(ii) $\int_{-\infty}^{\infty} K(x)dx = 1$

(iii) $\int_{-\infty}^{\infty} x^2 K(x)dx = \sigma^2 > 0$

(iv) $\int_{-\infty}^{\infty} xK(x)dx = 0$

Beberapa jenis fungsi kernel antara lain:

1. Kernel Uniform : K(x)=$\frac{1}{2}$; $|x| \leq 1$, 0 selainnya
2. Kernel Triangle : K(x) = (1-$|x|$) ; $|x| \leq 1$, 0 selainnya
3. Kernel Epanechnikov : K(x) = $\frac{3}{4}(1 - x^2)$; $|x| \leq 1$, 0 selainnya
4. Kernel Kuartik : K(x) = $\frac{15}{16}(1 - x^2)^2$; $|x| \leq 1$, 0 selainnya
5. Kernel Triweight : K(x) = $\frac{35}{32}(1 - x^2)^3$; $|x| \leq 1$, 0 selainnya
6. Kernel Cosinus : K(x) = $\frac{\pi}{4}\cos(\frac{\pi}{2}x)$; $|x| \leq 1$, 0 selainnya
7. Kernel Gaussian : K(x) = $\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$; $|x| \leq 1$, 0 selainnya

Estimator densitas kernel dari untuk fungsi densitas f(x) didefinisikan sebagai:

$$f_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \quad (8)$$

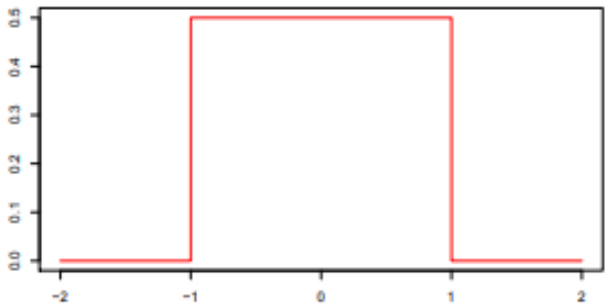
Dari persamaan (8) terlihat bahwa $f_h(x)$ tergantung pada fungsi kernel K dan parameter *h*. Bentuk bobot kernel ditentukan oleh fungsi kernel K, sedangkan ukuran bobotnya ditentukan oleh parameter pemulus *h* yang disebut *bandwidth* (Wand and Jones, 1995).

2.5 Kernel Uniform

Menurut Wand dan Jones (1995), kernel uniform adalah distribusi uniform untuk interval $[x^* - h, x^* + h]$, dimana h adalah bandwidth. Adapun fungsi densitas kernel uniform adalah sebagai berikut:

$$k_{x^*}(x) = \begin{cases} \frac{1}{2h} & x \in [x^* - h, x^* + h] \\ 0 & \text{Lainnya} \end{cases} \tag{9}$$

Bentuk kurva dari kernel uniform adalah:



Gambar 2.1 Kurva kernel uniform

2.6 Ukuran Kebaikan *Bandwidth* Optimal

Parameter *smoothing* yang digunakan untuk mengontrol kemulusan dari kurva yang diestimasi disebut dengan *bandwidth* (*h*). Proses pemilihan *bandwidth* yang sesuai (parameter *smoothing*) adalah bagian yang penting dari regresi nonparametrik. Permasalahan utama pada kernel *smoothing* bukan terletak pada pemilihan kernel tetapi pada pemilihan *bandwidth* (Hastie and Tibshirani, 1990).

Silverman (1986) memberikan lebar jendela untuk data melenceng dan *long-tailed* dengan rumus berikut:

$$h = 0.79(\text{IQR})n^{-1/5} \tag{10}$$

dengan,

h : *bandwidth*

IQR : *Inter Quartil Range* (simpangan antar kuartil)

n : banyaknya data

Kurva yang *under-smoothing* yaitu sangat kasar dan fluktuatif karena nilai *bandwidth* yang terlalu kecil. Sebaliknya, kurva yang *over-smoothing* yaitu sangat mulus karena nilai *bandwidth* yang terlalu lebar (Hardle, 1994). Sehingga diperlukan *bandwidth* yang optimal untuk menghasilkan kurva optimal. *Generalized Cross Validation* (GCV) merupakan salah satu metode untuk mendapatkan *h* yang optimal (Eubank, 1988), didefinisikan sebagai berikut:

$$GCV(h) = \frac{MSE}{(\frac{1}{n}TR(I-h(W)))^2} \tag{11}$$

dengan,

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{12}$$

2.7 Regresi Semiparametrik

Menurut Nanda (2016), regresi semiparametrik dapat digunakan apabila dalam suatu kasus terdiri dari komponen parametrik yang diketahui polanya dan komponen nonparametrik yang *smooth* tidak diketahui polanya.

Bentuk umum model regresi semiparametrik dapat dituliskan sebagai berikut:

$$y_i = \sum_{k=1}^{p} \beta_k x_{ik} + m(x_i) + \varepsilon_i \quad , i = 1, 2, \dots, n \, , k = 1, 2, \dots, p \qquad (13)$$

dengan,

y_i : variabel respon pengamatan ke-i

x_{ik} : variabel bebas parametrik ke-k pengamatan ke-i

x_i : variabel bebas nonparametrik pengamatan ke-i

β_k : parameter variable bebas x_k

m : komponen nonparametrik dengan fungsi regresi yang tidak diketahui

ε_i : galat ke-i

Persamaan (13) dapat juga ditulis menjadi $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{M}(\boldsymbol{X}) + \boldsymbol{\varepsilon}$. Menurut Ruppert, *et al*. (2003), adapun asumsi yang digunakan pada regresi semiparametrik adalah:

a. $\varepsilon \sim N(0, \sigma^2)$ dengan model acak dan saling bebas
b. Antara variabel prediktor adalah saling bebas (nonmultikolinieritas)
c. Sebagian bentuk kurva diketahui dan sebagian bentuk kurva tidak diketahui.

2.8 Estimasi Model Semiparametrik Kernel

a. *Partially Linear Model*

Model regesi semiparametrik pada persamaan (13), estimasi dilakukan secara parsial yaitu mengestimasi variabel parametrik dan variabel nonparametrik. Estimasi parameter dilakukan dengan menggunakan metode kuadrat terkecil

(MKT) pada model $\widehat{y_i} = \sum_{k=1}^{n} \beta_k x_{ik}$ dengan x adalah variabel parametrik, kemudian mengestimasi kurva regresi menggunakan kernel pada model $\widehat{y_i} = m(x_i)$ dengan x adalah variabel nonparametrik.

b. Estimasi Parameter Variabel Parametrik

Model estimasi regresi semiparametrik adalah sebagai berikut:

$$y_i = \sum_{k=1}^{p} \beta_k x_{ik} + \widehat{m}(x_i) + \varepsilon_i \qquad (14)$$

Persamaan (14) juga dapat ditulis menjadi $\boldsymbol{Y} = \boldsymbol{X\beta} + \widehat{\boldsymbol{M}}(\boldsymbol{x}) + \boldsymbol{\varepsilon}$. Dengan metode kuadrat terkecil diperoleh estimasi β dengan meminimumkan jumlah kuadrat galatnya, yaitu sebagai berikut:

$$\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} = [\boldsymbol{Y} - \boldsymbol{X\beta} - \widehat{\boldsymbol{M}}(\boldsymbol{x})]^T[\boldsymbol{Y} - \boldsymbol{X\beta} - \widehat{\boldsymbol{M}}(\boldsymbol{x})]$$

$$= [\boldsymbol{Y} - \boldsymbol{X\beta} - W(x)(\boldsymbol{Y} - \boldsymbol{X\beta})]^T[\boldsymbol{Y} - \boldsymbol{X\beta} - W(x)(\boldsymbol{Y} - \boldsymbol{X\beta})]$$

$$= [(\boldsymbol{Y} - W(x)\boldsymbol{Y}) - (\boldsymbol{X} - W(x)\boldsymbol{X})\boldsymbol{\beta}]^T[(\boldsymbol{Y} - W(x)\boldsymbol{Y}) - (\boldsymbol{X} - W(x)\boldsymbol{X})\boldsymbol{\beta}]$$

Misalkan $\widetilde{\boldsymbol{Y}} = \boldsymbol{Y} - W(x)\boldsymbol{Y}$ dan $\widetilde{\boldsymbol{X}} = \boldsymbol{X} - W(x)\boldsymbol{X}$, maka:

$$\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} = \left(\widetilde{\boldsymbol{Y}} - \widetilde{\boldsymbol{X}}\boldsymbol{\beta}\right)^T(\widetilde{\boldsymbol{Y}} - \widetilde{\boldsymbol{X}}\boldsymbol{\beta})$$

$$= \widetilde{\boldsymbol{Y}}^T\widetilde{\boldsymbol{Y}} - \boldsymbol{2\beta}^T\widetilde{\boldsymbol{X}}^T\widetilde{\boldsymbol{Y}} + \boldsymbol{\beta}^T\widetilde{\boldsymbol{X}}^T\widetilde{\boldsymbol{X}}\boldsymbol{\beta}$$

$\frac{\partial \boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}}{\partial \boldsymbol{\beta}} = 0$, maka diperoleh $-\boldsymbol{2}\widetilde{\boldsymbol{X}}^T\widetilde{\boldsymbol{Y}} + \boldsymbol{2}\widetilde{\boldsymbol{X}}^T\widetilde{\boldsymbol{X}}\widehat{\boldsymbol{\beta}} = \boldsymbol{0}$

$$\widetilde{\boldsymbol{X}}^T\widetilde{\boldsymbol{Y}} = \widetilde{\boldsymbol{X}}^T\widetilde{\boldsymbol{X}}\widehat{\boldsymbol{\beta}}$$

$\frac{\partial^2 \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} = \mathbf{2}\widetilde{\boldsymbol{X}}^T \widetilde{\boldsymbol{X}}$ merupakan matriks semidefinit positif. Jika $\mathbf{2}\widetilde{\boldsymbol{X}}^T \widetilde{\boldsymbol{X}}$ adalah matriks definit postif, maka $\widehat{\boldsymbol{\beta}}$ merupakan estimator kuadrat terkecil untuk β dengan:

$$\widehat{\boldsymbol{\beta}} = (\widetilde{\boldsymbol{X}}^T \widetilde{\boldsymbol{X}})^{-1} \widetilde{\boldsymbol{X}}^T \widetilde{\boldsymbol{Y}} \tag{15}$$

Berdasarkan uraian di atas, jika $\widehat{y}_i$ adalah variabel penduga untuk y_i diperoleh model dugaan semiparametrik kernel yaitu $\widehat{y}_i = \sum_{k=1}^{p} \beta_k x_{ik} + \widehat{m}(x_i)$ dengan $\widehat{m}(x_i) = \sum_{j=1}^{n} W_j(x_i)(y_j - \sum_{k=1}^{p} \widehat{\beta}_k x_{ik})$ dan $\widehat{\boldsymbol{\beta}} = (\widetilde{\boldsymbol{X}}^T \widetilde{\boldsymbol{X}})^{-1} \widetilde{\boldsymbol{X}}^T \widetilde{\boldsymbol{Y}}$.

c. Estimasi Kurva Regresi Variabel Nonparametrik

Pada model regresi semiparametrik pada persamaan (13), m dan β merupakan fungsi nonparametrik dan parameter yang akan diestimasi dari data. Estimasi m menggunakan estimator kernel. Adapun penduga fungsi regresi semiparametrik adalah sebagai berikut:

$$\widehat{m}(x) = \sum_{i=1}^{n} w_{hi}(x) y_i \tag{16}$$

dengan,

$$w_{hi}(x) = \frac{K_h(x - x_i)}{\sum_{i=1}^{n} K_h(x - x_i)}$$

$$= \frac{\frac{1}{h} K(\frac{x - x_i}{h})}{\frac{1}{h} \sum_{i=1}^{n} K(\frac{x - x_i}{h})}$$

$$= \frac{K(\frac{x - x_i}{h})}{\sum_{i=1}^{n} K(\frac{x - x_i}{h})} \tag{17}$$

Penduga (17) diusulkan oleh Nadaraya dan Watson, sehingga penduga ini sering disebut penduga Nadaraya-Watson (Hardle, 1994).

Misalkan jika terdapat *n* observasi dengan *p* variable *x*, maka persamaan (6) dapat ditulis sebagai berikut:

$$y_i - \sum_{k=1}^{p} \beta_k x_{ik} = m(x) + \varepsilon_i \tag{18}$$

Jika $y_i{}^{*} = y_i - \sum_{k=1}^{p} \beta_k x_{ik}$ maka persamaan estimator Nadaraya-Watson seperti regresi nonparametrik kernel diperoleh sebagai berikut:

$$\widehat{m}(x) = \frac{K(\frac{x-x_i}{h})y_i{}^{*}}{\sum_{i=1}^{n} K(\frac{x-x_i}{h})} = \sum_{i=1}^{n} W_i(x)(y_i - \sum_{k=1}^{p} \beta_k x_{ik}) \tag{19}$$

2.9 Ukuran Kebaikan Model

Kebaikan suatu model dapat dilihat dari tingkat kesalahannya, semakin kecil tingkat kesalahan semakin baik estimasinya. Terdapat beberapa kriteria untuk menentukan tingkat kesalahan antara lain *Mean Square Error (*MSE), *Root Mean Square Error (*RMSE), *Mean Absolute Error* (MAE), dan *Mean Absolute Percentage Error* (MAPE). Dapat juga dengan mengamati nilai koefisien determinasi didefinisikan sebagai:

$$R^2 = \frac{JKR}{JKT} \tag{20}$$

dengan $0 \leq R^2 \leq 1$, dimana makin dekat R^2 dengan 1 semakin baik estimasinya, dan sebaliknya, makin dekat R^2 dengan 0 semakin jelek estimasinya.

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilakukan pada semester ganjil 2018-2019 dan 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 dan data sekunder yaitu data rekam medis pasien DBD mulai bulan Januari-Agustus 2018 yang didapat dari Rumah Sakit Umum Dr. H Abdul Moeleoek Provinsi Lampung. Variabel-variabel yang digunakan pada penelitian ini terdiri dari variabel respon dan variabel prediktor. Variabel respon (Y) pada penelitian ini adalah lama kesembuhan pasien DBD. Sedangkan variabel prediktornya (X) yaitu suhu tubuh (X_1), kadar hemoglobin (X_2), dan kadar hematokrit (X_3) pasien DBD.

3.3 Metode Penelitian

Dalam penelitian ini model regresi semiparametrik diduga menggunakan penduga (*estimator*) kernel dengan fungsi kernel uniform. Langkah-langkah yang dilakukan adalah:

1. Mendeksripsikan data asli, lama kesembuhan pasien DBD dan faktor yang diduga mempengaruhi yaitu data suhu tubuh, kadar hemoglobin, dan kadar hematokrit. Begitupun dengan data bangkitan, mendeskripsikan data X_1, X_2, dan X_3.
2. Melakukan simulasi data dengan langkah-langkah sebagai berikut:
   a. Dibangkitkan variabel bebas (X_i) yang mengikuti data asli sebanyak p = 3, n = 30 dengan $\beta_1 = 0.8666508$, $\beta_2 = 1.1225459$ dengan

      $X_1 \sim Normal(37.6167, 0.4939)$

      $X_2 \sim Normal(31.3, 3.75)$

      $X_3 \sim Weibull(10.9, 41.65)$

   b. Mendapat nilai Y_i dengan model persamaan :

      $Y_i = \beta_1 X_1 + \beta_2 X_2 + m(X_i) + \varepsilon_i$
3. Membuat diagram pencar pada masing-masing variabel prediktor dengan variabel respon untuk menentukan variabel parametrik dan variabel nonparametrik.
4. Mengestimasi parameter regresi pada variabel parametrik menggunakan metode kuadrat terkecil pada persamaan (19).
5. Melakukan uji asumsi klasik terhadap galat pada variabel parametrik.

6. Menghitung nilai *bandwidth* menggunakan rumus lebar jendela Silverman pada persamaan (11).
7. Menentukan parameter pemulus (*bandwidth*) optimal menggunakan fungsi kernel uniform dan ditentukan dengan GCV yang memenuhi kriteria minimum.
8. Mengestimasi kurva regresi nonparametrik menggunakan *bandwidth* optimal.
9. Mendapatkan estimasi model regresi semiparametrik berdasarkan hasil estimasi parameter regresi parametrik dan estimasi kurva regresi nonparametrik.
10. Menginterpretasikan model yang diperoleh.

V. KESIMPULAN

Berdasarkan hasil penelitian diperoleh kesimpulan bahwa estimasi model regresi semiparametrik dengan fungsi kernel pada data rekam medis pasien DBD di RSUD Abdoel Moeloek Provinsi Lampung menghasilkan model sebagai berikut:

$$\hat{y}_i = 0.8671058X_1 + 1.1225459X_2 + \widehat{m}_h(X_3)$$

Sementara, hasil penelitian estimasi model regresi semiparametrik dengan fungsi kernel pada data simulasi yang mengitkuti sebaran data asli menghasilkan model sebagai berikut:

$$\widehat{y_i} = 0.866605X_1 + 1.122494X_2 + \widehat{m}_h(X_3)$$

DAFTAR PUSTAKA

Anselin, L. 1988. *Spatial Econometrics: Methods and Models.* Kluwer Academic Publisher, Dordecht.

Corder, G.W. and Foreman, D.I. 2009. *Nonparametrics for Non-Statisticians*. John Wiley and Sons, New York.

Draper, N. dan Smith, H. 1992. *Analisis Regresi Terapan.* Ed. ke-2. Gramedia Pustaka, Jakarta.

Eubank, R. 1998. *Spline Smoothing and Nonparametric Regression*. Marcel Dekker, New York.

Fitriani, A., *et al.* 2015. Estimasi Model Regresi Semiparametrik Menggunakan Estimator Kernel Uniform (Studi Kasus: Pasien DBD di RS Puri Raharja). *E-Journal Matematika*. **4**(4):176-180.

Hastie, T and Tibshirani, R. 1990. *Generalized Additive Models.* Chapman and Hall, New York.

Hosmer, D.W. and Lemeshow, S. 2000. *Applied Logistic Regression.* Ed. ke-2. John Wiley and Sons, New York.

Kutner, M.H., *et. al.* 2004. *Applied Linear Regression Models.* Ed. ke-4. McGraw-Hill Company, Inc. New York.

Marlin, A.F. 2016. Penduga Kurva Regresi Nonparametrik Linear dan Nonlinear dengan Metode Prestley-Chao, Nadaraya-Watson dan Metode Fourier. Skripsi. Jurusan Matematika FMIPA Universitas Lampung, Lampung.

Nanda, D.A., *et al*. 2016. Analisis Pengaruh Jumlah Uang Beredar dan Nilai Tukar Rupiah Terhadap Indeks Harga Saham Gabungan Menggunakan Pemodelan Regresi Semiparametrik Kernel. *Jurnal Gaussian*. **5**(3): 373-382.

Riduwan, A. 2006. *Rumus dan Data dalam Aplikasi Statistika.* Alfabeta, Bandung.

Ruppert, *et al*. 2003. *Semiparametric Regression*. Cambridge University Press, New York.

Santoso, S. 2012. *Statistik Parametrik.* Gramedia Pustaka, Jakarta.

Silverman, B.W. 1986. *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, New York.

Wand, M.P. and Jones, M.C. 1995. *Kernel Smoothing*. Chapman and Hall, New York.

Wu, H. and Zhang, J.T. 2006. *Nonparametric Regression Methods for Longitudinal Data Analysis*. John Wiley and Sons, New Jersey.