

**PERBANDINGAN ESTIMATOR KERNEL PADA *BANDWIDTH* OPTIMUM
UNTUK DATA PENERIMAAN PAJAK KENDARAAN BERMOTOR (PKB)
PROVINSI LAMPUNG**

(Skripsi)

Oleh

DONY CHERISTIAN VENESIA



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

ABSTRACT

COMPARISON OF KERNEL ESTIMATOR ON OPTIMUM *BANDWIDTH* FOR MOTORIZED VEHICLE TAX (MVT) REVENUE DATA IN LAMPUNG PROVINCE

By

DONY CHERISTIAN VENESIA

Regression analysis is used to build a functional model of several variables to explain or predict a natural phenomenon based on other phenomena. The pattern of motorized vehicle tax distribution is assumed not to follow the patterns of certain distribution. Therefore, the nonparametric regression is used to estimate the distribution. In nonparametric regression, smoothing techniques are used to estimate the distribution of data, one of which is the Kernel estimator. There are four Kernel functions, namely Biweight, Triangular, Epanechnikov, and Rectangular. The most important in the Kernel estimator is to determine the optimum bandwidth based on the minimum GCV value. The results of the data analysis show that estimating motorized vehicle tax revenue data using the four Kernel estimators will produce a different optimum bandwidth. The optimum bandwidth for the Kernel Biweight estimator is at 5,1 with a MSE value of $1,640395 \times 10^{19}$, the optimum bandwidth for the Kernel Triangular estimator is at 4,8 with a MSE value of $1,536755 \times 10^{19}$, the optimum bandwidth for the Kernel Epanechnikov estimator is at a value of 4,4 with a MSE value of $1,705738 \times 10^{19}$, while the optimum bandwidth for the Rectangular Kernel estimator is at value 2 with a MSE value of $1,606819 \times 10^{19}$. From the results of the comparison above, it can be concluded that the best estimator in the motorized vehicle tax revenue data is the Triangular Kernel estimator, because it has the smallest MSE value.

Keywords: Nonparametric Regression, Kernel Estimator, Kernel Function, Bandwidth, Motorized Vehicle Tax

ABSTRAK

PERBANDINGAN ESTIMATOR KERNEL PADA *BANDWIDTH* OPTIMUM UNTUK DATA PENERIMAAN PAJAK KENDARAAN BERMOTOR (PKB) PROVINSI LAMPUNG

Oleh

DONY CHERISTIAN VENESIA

Analisis regresi merupakan ilmu statistik yang mempelajari bagaimana membangun sebuah model fungsional dari beberapa variabel sehingga dapat menjelaskan atau meramalkan suatu fenomena alami berdasarkan fenomena-fenomena yang lain. Pola sebaran data penerimaan pajak kendaraan bermotor diasumsikan tidak mengikuti pola sebaran data tertentu. Oleh karena itu, untuk mengestimasi sebaran data digunakan regresi nonparametrik. Dalam regresi nonparametrik, untuk mengestimasi sebaran data digunakan teknik *smoothing*, salah satunya yaitu estimator Kernel. Terdapat empat fungsi Kernel, yaitu *Biweight*, *Triangular*, *Epanechnikov*, dan *Rectangular*. Hal terpenting dalam estimator Kernel adalah penentuan *bandwidth* optimum didasarkan pada nilai GCV minimum. Hasil analisis data menunjukkan bahwa mengestimasi data penerimaan pajak kendaraan bermotor menggunakan keempat estimator Kernel akan menghasilkan *bandwidth* optimum yang berbeda. *Bandwidth* optimum untuk estimator *Kernel Biweight* berada pada nilai 5,1 dengan nilai MSE sebesar $1,640395 \times 10^{19}$, *bandwidth* optimum untuk estimator *Kernel Triangular* berada pada nilai 4,8 dengan nilai MSE sebesar $1,536755 \times 10^{19}$, *bandwidth* optimum untuk estimator *Kernel Epanechnikov* berada pada nilai 4,4 dengan nilai MSE sebesar $1,705738 \times 10^{19}$, sedangkan *bandwidth* optimum untuk estimator *Kernel Rectangular* berada pada nilai 2 dengan nilai MSE sebesar $1,606819 \times 10^{19}$. Dari hasil perbandingan nilai MSE untuk keempat estimator, disimpulkan bahwa estimator terbaik pada data penerimaan pajak kendaraan bermotor adalah estimator *Kernel Triangular*, karena memiliki nilai MSE terkecil.

Kata Kunci: Regresi Nonparametrik, Estimator Kernel, Fungsi Kernel, *Bandwidth*, Pajak Kendaraan Bermotor

**PERBANDINGAN ESTIMATOR KERNEL PADA *BANDWIDTH* OPTIMUM
UNTUK DATA PENERIMAAN PAJAK KENDARAAN BERMOTOR (PKB)
PROVINSI LAMPUNG**

Oleh

Dony Pheristian Venesia

Skripsi

Sebagai Salah Satu Syarat untuk Memperoleh Gelar
SARJANA MATEMATIKA

pada

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung



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

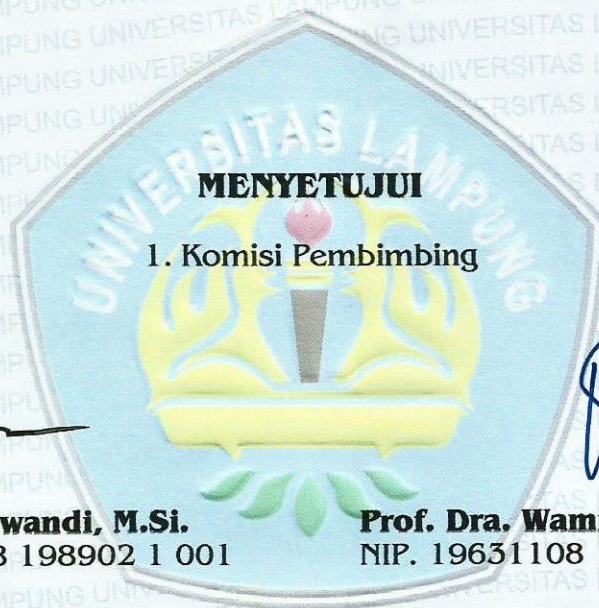
Judul Skripsi : **PERBANDINGAN ESTIMATOR KERNEL
PADA BANDWIDTH OPTIMUM UNTUK DATA
PENERIMAAN PAJAK KENDARAAN
BERMOTOR (PKB) PROVINSI LAMPUNG**

Nama Mahasiswa : **Dony Cheristian Venesia**

Nomor Pokok Mahasiswa : 1517031023

Jurusan : Matematika

Fakultas : Matematika dan Ilmu Pengetahuan Alam




Drs. Rudi Ruswandi, M.Si.
NIP. 19560208 198902 1 001


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

2. Ketua Jurusan Matematika


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

MENGESAHKAN

1. Tim Penguji

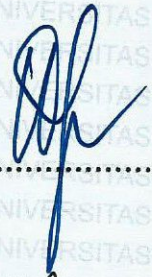
Ketua

: Drs. Rudi Ruswandi, M.Si.



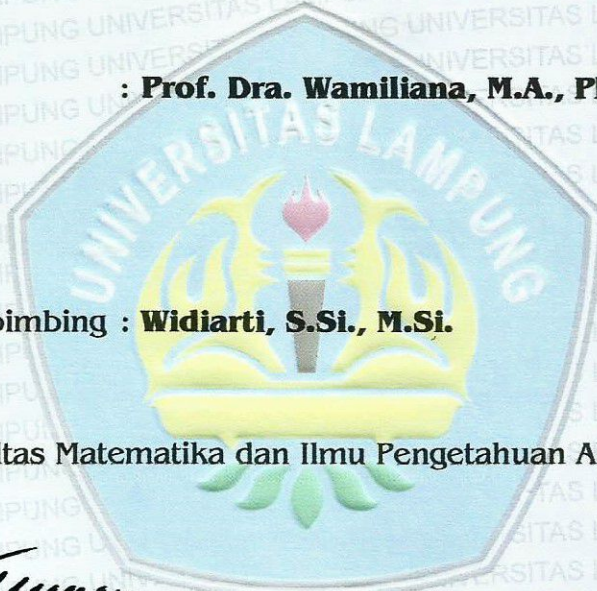
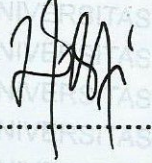
Sekretaris

: Prof. Dra. Wamiliana, M.A., Ph.D.



Penguji

Bukan Pembimbing : Widiarti, S.Si., M.Si.



2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



Drs. Suratman, M.Sc.

NIP. 19640604 199003 1 002

Tanggal Lulus Ujian Skripsi : 16 Desember 2019

PERNYATAAN SKRIPSI MAHASISWA

Saya yang bertanda tangan di bawah ini:

Nama : **Dony Cheristian Venesia**

Nomor Pokok Mahasiswa : **1517031023**

Jurusan : **Matematika**

Judul Skripsi : **Perbandingan Estimator Kernel pada *Bandwidth* Optimum untuk Data Penerimaan Pajak Kendaraan Bermotor (PKB) Provinsi Lampung**

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, 16 Desember 2019

Yang Menyatakan



Dony Cheristian Venesia
NPM. 1517031023

RIWAYAT HIDUP

Penulis bernama lengkap Dony Cheristian Venesia, lahir di Sungai Liat pada tanggal 15 Desember 1997. Penulis merupakan anak keempat dari empat bersaudara yang lahir dari pasangan Bapak Lianarfin dan Ibu Halimah.

Penulis menempuh pendidikan sekolah dasar di SD Negeri 2 Sukadana Pasar Kabupaten Lampung Timur pada tahun 2003-2009, kemudian melanjutkan ke jenjang sekolah menengah pertama di SMP Negeri 1 Sukadana Kabupaten Lampung Timur pada tahun 2009-2012, dan jenjang sekolah menengah atas di SMA Negeri 1 Sukadana Kabupaten Lampung Timur pada tahun 2012-2015. Pada tahun 2015, Penulis diterima sebagai mahasiswa S1 di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung melalui jalur Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN) dan merupakan salah satu mahasiswa penerima Beasiswa Bidik Misi.

Pada tahun 2018, Penulis melaksanakan Praktik Kerja Lapangan (PKL) di Kantor Badan Pendapatan Daerah Provinsi Lampung, sebagai bentuk aplikasi bidang ilmu di dunia kerja. Pada tahun yang sama Penulis juga melaksanakan Kuliah Kerja Nyata (KKN) di Desa Candra Kencana, Kecamatan Tulang Bawang Tengah, Kabupaten Tulang Bawang Barat, sebagai bentuk pengabdian kepada masyarakat.

KATA INSPIRASI

“Sesungguhnya sesudah kesulitan itu ada kemudahan. Maka apabila kamu telah selesai (dari suatu urusan), kerjakanlah dengan sungguh-sungguh (urusan yang lainnya). Dan hanya kepada Tuhanmu lah hendaknya kamu berharap”

(QS. Al-Insyirah : 6-8)

“Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya. Dia mendapat (pahala) dari (kebajikan) yang dikerjakannya dan dia mendapat (siksa) dari (kejahatan) yang diperbuatnya”

(QS. Al-Baqarah : 286)

“Maka nikmat Tuhanmu yang manakah yang kamu dustakan?”

(Q.S. Ar-Rahmaan: 13)

“Bayangkan kita hidup saat ini dalam keadaan telah mati, agar kita bisa berfikir dan merasakan betapa menyesalnya jika tidak menjauhi apa yang dilarang dan tidak menjalani apa yang diperintah oleh Allah Subhanahu Wa Ta'ala dan sunnah Rasulullah Shallallahu

‘Alaihi Wa Sallam saat kita masih hidup di dunia”

(Dony Christian Venesia)

PERSEMBAHAN

Alhamdulillah, puji syukur kehadiran Allah SWT yang Maha Pengasih lagi Maha Penyayang. Dengan segala kerendahan hati penulis persembahkan skripsi ini kepada:

Kedua Orang Tua Tercinta

yang selalu mencurahkan seluruh tenaga dan pikiran untuk keberhasilan Penulis dalam menuntut ilmu, memberikan dukungan, bimbingan, nasihat, dan doa yang terus mengalir untuk kesuksesan Penulis.

Kakek-kakak Tersayang

yang senantiasa memberikan dukungan, motivasi, keceriaan, dan selalu memanjatkan doa untuk keberhasilan dan kesuksesan Penulis.

Dosen Pembimbing dan Pembahas

atas bimbingan dan motivasi dalam menyelesaikan skripsi.

Sahabat-sahabat Terbaik

yang selalu memberikan kebahagiaan serta keceriaan, terima kasih atas semua kenangan dan pengalaman terindah yang telah kita lalui bersama. Semoga dapat menjadi pembelajaran berharga untuk kehidupan kita yang lebih baik,

Almamater Tercinta, Universitas Lampung

SANWACANA

Puji dan syukur penulis panjatkan kehadiran Allah SWT, yang telah memberikan rahmad dan hidayah-Nya kepada penulis, sehingga penulis dapat menyelesaikan skripsi yang berjudul “Perbandingan Estimator Kernel pada *Bandwidth* Optimum untuk Data Penerimaan Pajak Kendaraan Bermotor (PKB) Provinsi Lampung”.

Penulis menyadari bahwa skripsi ini tidak akan terwujud tanpa adanya bantuan, bimbingan, dan doa dari berbagai pihak sehingga skripsi ini dapat diselesaikan.

Dengan segala kerendahan hati penulis mengucapkan terima kasih kepada:

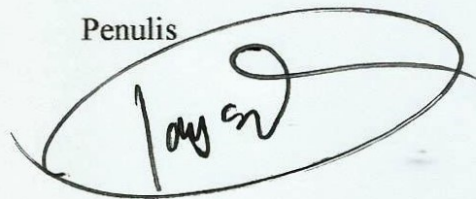
1. Bapak Drs. Rudi Ruswandi, M.Si., selaku dosen Pembimbing I serta selaku pembimbing akademik yang memberikan motivasi, bimbingan, pengarahan, dan saran kepada penulis dalam menyelesaikan skripsi serta telah banyak membimbing selama proses perkuliahan.
2. Ibu Prof. Dra. Wamiliana, M.A., Ph.D., selaku Pembimbing II serta selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung atas kesediaan memberikan bimbingan, saran, dan kritik dalam proses penyelesaian skripsi ini.
3. Ibu Widiarti, S.Si., M.Si., selaku Pembahas skripsi yang telah memberikan evaluasi dan saran bagi perbaikan skripsi penulis.
4. Bapak Drs. Suratman, M.Sc., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

5. Seluruh dosen, staf, dan karyawan Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
6. Ayah, Ibu, Kakak-kakak, dan seluruh keluarga besar yang selalu memberi dukungan, motivasi, semangat, dan perhatian kepada penulis sehingga dapat menyelesaikan skripsi ini.
7. Sahabat-sahabat terbaik yang selalu memberikan semangat kepada penulis dalam menyusun skripsi ini yang tidak dapat disebutkan satu persatu.
8. Teman-teman mahasiswa Jurusan Matematika angkatan 2015 dan keluarga besar Matematika FMIPA Unila.
9. Seluruh pihak yang telah membantu dan terlibat dalam menyelesaikan skripsi ini.

Penulis menyadari bahwa dalam penulisan skripsi ini masih jauh dari kesempurnaan, oleh karena itu kritik dan saran sangat penulis harapkan. Semoga skripsi ini dapat bermanfaat bagi penulis dan bagi para pembaca.

Bandar Lampung, 16 Desember 2019

Penulis

A handwritten signature in black ink, enclosed within a hand-drawn oval. The signature appears to be 'Dony Christian Venesia' written in a cursive style.

Dony Cheristian Venesia

DAFTAR ISI

| | Halaman |
|---|---------|
| DAFTAR ISI | xiii |
| DAFTAR TABEL | xv |
| DAFTAR GAMBAR | xvi |
| | |
| I. PENDAHULUAN | |
| 1.1 Latar Belakang dan Masalah..... | 1 |
| 1.2 Tujuan Penelitian | 4 |
| 1.3 Manfaat Penelitian | 5 |
| | |
| II. TINJAUAN PUSTAKA | |
| 2.1 Regresi Nonparametrik | 6 |
| 2.2 Estimator Kernel | 7 |
| 2.3 Fungsi Kernel..... | 9 |
| 2.3.1 Definisi Fungsi Kernel | 9 |
| 2.3.2 Macam-Macam Fungsi Kernel | 10 |
| 2.4 Regresi Kernel | 11 |
| 2.4.1 Estimator <i>Kernel Biweight</i> dengan Metode Nadaraya-Watson..... | 15 |
| 2.4.2 Estimator <i>Kernel Triangular</i> dengan Metode Nadaraya-Watson .. | 16 |
| 2.4.3 Estimator <i>Kernel Epanechnikov</i> dengan Metode Nadaraya-Watson | 16 |
| 2.4.4 Estimator <i>Kernel Rectangular</i> dengan Metode Nadaraya-Watson | 17 |
| 2.5 Pemilihan <i>Bandwidth</i> Optimum..... | 18 |
| 2.6 Pajak Kendaraan Bermotor | 20 |

III. METODOLOGI PENELITIAN

| | |
|--------------------------------------|----|
| 3.1 Waktu dan Tempat Penelitian..... | 22 |
| 3.2 Data Penelitian..... | 22 |
| 3.3 Metode Penelitian | 23 |

IV. HASIL DAN PEMBAHASAN

| | |
|--|----|
| 4.1 Data..... | 27 |
| 4.2 Analisis Data..... | 28 |
| 4.2.1 Estimator <i>Kernel Biweight</i> dengan Metode Nadaraya-Watson..... | 28 |
| 4.2.2 Estimator <i>Kernel Triangular</i> dengan Metode Nadaraya-Watson .. | 33 |
| 4.2.3 Estimator <i>Kernel Epanechnikov</i> dengan Metode Nadaraya-Watson | 39 |
| 4.2.4 Estimator <i>Kernel Rectangular</i> dengan Metode Nadaraya-Watson | 44 |
| 4.2.5 Perbandingan Estimator <i>Kernel Biweight, Triangular, Epanechnikov, dan Rectangular</i> | 50 |

V. KESIMPULAN DAN SARAN

| | |
|----------------------|----|
| 5.1 Kesimpulan | 52 |
| 5.2 Saran | 52 |

DAFTAR PUSTAKA

LAMPIRAN

DAFTAR TABEL

| Tabel | Halaman |
|--|---------|
| 1. Macam-Macam Fungsi Kernel..... | 10 |
| 2. Data Penerimaan Pajak Kendaraan Bermotor..... | 22 |
| 3. Perbandingan Nilai MSE (<i>Mean Square Error</i>) | 50 |

DAFTAR GAMBAR

| Gambar | Halaman |
|--|---------|
| 1. Diagram Pencar Penerimaan Pajak Kendaraan Bermotor Tahun 2015 sampai dengan 2017 | 27 |
| 2. Estimasi Kurva Regresi <i>Kernel Biweight</i> dengan <i>Bandwidth</i> Sebesar 0,1 | 29 |
| 3. Estimasi Kurva Regresi <i>Kernel Biweight</i> dengan <i>Bandwidth</i> Sebesar 10 | 30 |
| 4. Estimasi Kurva Regresi <i>Kernel Biweight</i> dengan <i>Bandwidth</i> Optimum | 33 |
| 5. Estimasi Kurva Regresi <i>Kernel Triangular</i> dengan <i>Bandwidth</i> Sebesar 0,1 .. | 35 |
| 6. Estimasi Kurva Regresi <i>Kernel Triangular</i> dengan <i>Bandwidth</i> Sebesar 10 ... | 36 |
| 7. Estimasi Kurva Regresi <i>Kernel Triangular</i> dengan <i>Bandwidth</i> Optimum | 38 |
| 8. Estimasi Kurva Regresi <i>Kernel Epanechnikov</i> dengan <i>Bandwidth</i> Sebesar 0,1 | 40 |
| 9. Estimasi Kurva Regresi <i>Kernel Epanechnikov</i> dengan <i>Bandwidth</i> Sebesar 10 | 41 |
| 10. Estimasi Kurva Regresi <i>Kernel Epanechnikov</i> dengan <i>Bandwidth</i> Optimum | 44 |
| 11. Estimasi Kurva Regresi <i>Kernel Rectangular</i> dengan <i>Bandwidth</i> Sebesar 0,1 | 46 |
| 12. Estimasi Kurva Regresi <i>Kernel Rectangular</i> dengan <i>Bandwidth</i> Sebesar 10. | 47 |
| 13. Estimasi Kurva Regresi <i>Kernel Rectangular</i> dengan <i>Bandwidth</i> Optimum... | 49 |

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Pajak merupakan salah satu sumber penerimaan negara yang sangat penting untuk membangun dan memperbaiki infrastruktur maupun meningkatkan perekonomian negara. Setiap tahun pemerintah berusaha untuk memaksimalkan penerimaan pajak untuk membiayai pengeluaran negara karena semakin tinggi tingkat penerimaan pajak maka semakin tinggi kemampuan negara membiayai pembangunan. Sebaliknya, jika semakin kecil penerimaan pajak maka semakin rendah kemampuan negara dalam hal mewujudkan pembangunan negara.

Salah satu jenis pajak yang mempengaruhi penerimaan negara adalah Pajak Kendaraan Bermotor (PKB). Dasar pengenaan pajak kendaraan bermotor adalah nilai jual kendaraan bermotor dan bobot yang mencerminkan kadar kerusakan jalan dan pencemaran lingkungan akibat penggunaan kendaraan bermotor tersebut. Nilai jual kendaraan bermotor sesuai dengan harga pasar kendaraan bermotor, jenis kendaraan bermotor, merk kendaraan bermotor, tahun pembuatan kendaraan bermotor, berat total kendaraan bermotor, serta dokumen impor jenis kendaraan tertentu.

Prosedur perhitungan pajak kendaraan bermotor yang baik merupakan salah satu aspek pengelolaan pendapatan negara yang sangat mendukung tercapainya tujuan yang ditetapkan oleh pemerintah. Dengan demikian, perlu dilakukan estimasi penerimaan pajak kendaraan bermotor secara baik sehingga dapat diketahui berapa besar penerimaan yang nantinya dapat dikelola secara baik pula.

Pola sebaran pada data pajak kendaraan bermotor diasumsikan tidak mengikuti pola sebaran data normal sehingga pada penelitian ini digunakan metode nonparametrik untuk mendekati pola sebaran data pajak kendaraan bermotor. Estimasi fungsi regresi nonparametrik dilakukan berdasarkan data pengamatan dengan menggunakan teknik *smoothing* (Eubank, 1999). Terdapat beberapa teknik *smoothing* dalam regresi nonparametrik antara lain histogram, estimator kernel, deret orthogonal, estimator spline, k-NN, deret Fourier, dan wavelet.

Dalam penelitian ini, estimator yang digunakan adalah estimator kernel karena estimator ini mempunyai bentuk yang fleksibel dan secara matematik mudah dikerjakan serta mempunyai rata-rata kekonvergenan yang relatif cepat (Hardle, 1990). Menurut Eubank (1999) pada dasarnya estimator kernel sama dengan estimator linier lainnya hanya saja metode kernel lebih khusus dalam penggunaan metode *bandwidth*.

Bandwidth dinotasikan dengan h adalah konstanta positif untuk menentukan kemulusan dari kurva taksiran regresi. Dengan memilih *bandwidth* yang mendekati 0 maka taksiran yang didapatkan tidak mulus bahkan akan menonjolkan plot datanya saja.

Pada regresi kernel pemilihan *bandwidth* jauh lebih penting daripada pemilihan fungsi kernel. Jika *bandwidth* yang dipilih terlalu kecil maka akan menghasilkan estimasi kurva kurang mulus (*under-smoothing*), sebaliknya jika *bandwidth* terlalu besar maka akan menghasilkan estimasi kurva sangat mulus (*over-smoothing*) yang tidak sesuai dengan pola sebaran data. Sehingga harus dipilih nilai *bandwidth* yang optimum agar dihasilkan estimasi terbaik. Terdapat beberapa metode yang digunakan dalam pemilihan *bandwidth* optimum, salah satunya adalah menggunakan kriteria *Generalized Cross Validation* (GCV) (Galub, dkk., 1979). *Bandwidth* optimum dipilih berdasarkan kriteria nilai GCV yang minimum. Selain itu, menurut Komang dan Gusti (2012) kebaikan suatu estimator dapat dilihat dari tingkat kesalahannya. Semakin kecil tingkat kesalahannya semakin baik estimasinya. Terdapat beberapa kriteria untuk menentukan estimator terbaik dalam model regresi nonparametrik, yaitu *Mean Square Error* (MSE), *Root Mean Square Error* (RMSE), dan *Mean Absolute Deviation* (MAD).

Salah satu teknik regresi nonparametrik yang sering digunakan untuk menaksir fungsi regresi $m(x)$ adalah dengan menggunakan estimator Nadaraya-Watson. Estimator ini diperoleh dengan menggunakan metode penaksiran fungsi densitas kernel. Ada beberapa jenis fungsi kernel, antara lain *Kernel Biweight*, *Triangular*, *Epanechnikov*, dan *Rectangular* (Hardle, 1990).

Berdasarkan uraian tersebut maka penulis mengambil judul “Perbandingan Estimator Kernel pada *Bandwidth* Optimum untuk Data Penerimaan Pajak Kendaraan Bermotor (PKB) Provinsi Lampung”. Penelitian ini menggunakan

data yang berasal dari Kantor Badan Pendapatan Daerah (BAPENDA) Provinsi Lampung yaitu penerimaan pajak kendaraan bermotor dan menggunakan fungsi *Kernel Biweight*, *Triangular*, *Epanechnikov*, dan *Rectangular* untuk mendekati pola sebaran data penerimaan pajak kendaraan bermotor. Untuk mengetahui estimasi sebaran data penerimaan pajak kendaraan bermotor serta estimator kernel terbaik dengan menggunakan estimator Nadaraya-Watson digunakan *software RStudio Version 1.1.419*.

1.2 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah:

1. Mengetahui *bandwidth* optimum untuk fungsi *Kernel Biweight*, *Triangular*, *Epanechnikov*, dan *Rectangular* berdasarkan kriteria *Generalized Cross Validation* (GCV) yang minimum.
2. Mengetahui estimator terbaik menggunakan metode Nadaraya-Watson antara estimator *Kernel Biweight*, *Triangular*, *Epanechnikov*, dan *Rectangular* pada data penerimaan pajak kendaraan bermotor.

1.3 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah:

1. Sebagai media untuk mengaplikasikan dan mengembangkan pengetahuan statistik, keterampilan riset dan keterampilan pengolahan data khususnya mengenai estimator kernel dalam mengestimasi sebaran data penerimaan pajak kendaraan bermotor.
2. Memperoleh kontribusi pemikiran dan informasi untuk pengembangan lembaga dan sumber daya yang ada agar dapat diketahui berapa besar penerimaan yang nantinya dapat dikelola secara baik.

II. TINJAUAN PUSTAKA

2.1 Regresi Nonparametrik

Salah satu teknik statistik yang sering digunakan dalam ilmu dan pengetahuan terapan adalah analisis regresi. Analisis regresi merupakan salah satu ilmu statistik yang mempelajari bagaimana membangun sebuah model fungsional dari data sehingga dapat menjelaskan atau meramalkan suatu fenomena alami (variabel Y) atas dasar fenomena yang lain (variabel X).

Pada analisis regresi, digunakan dua model pendekatan yaitu pendekatan parametrik dan nonparametrik. Model pendekatan parametrik, bentuk kurva regresi diasumsikan mengandung parameter tertentu sehingga untuk memperoleh estimator kurva regresi dilakukan dengan mengestimasi parameter tersebut. Sedangkan model pendekatan nonparametrik tidak mensyaratkan bentuk sebaran parameter populasi sehingga dapat digunakan pada data yang memiliki sebaran normal atau tidak. Pendekatan nonparametrik merupakan metode statistik yang dapat digunakan dengan mengabaikan asumsi-asumsi yang melandasi penggunaan metode statistik parametrik, terutama yang berkaitan dengan distribusi normal (Eubank, 1999).

Model regresi nonparametrik secara matematis dapat ditulis:

$$Y = m(x) + \varepsilon \quad (2.1.1)$$

Keterangan:

Y = variabel respon

$m(x)$ = fungsi regresi nonparametrik yang memuat variabel prediktor

ε = *error* (galat) *absolute*, dirumuskan dengan $\varepsilon = Y - m(x)$. Jika tanda *error* (positif atau negatif) tidak dipertimbangkan maka diperoleh *error* mutlak sebagai berikut: $|\varepsilon| = |Y - m(x)|$

Menurut Eubank (1999) regresi nonparametrik adalah suatu teknik analisis data yang menjelaskan hubungan antara variabel prediktor dengan variabel respon yang tidak diketahui bentuk fungsinya. Tetapi diasumsikan mulus (*smooth*) dalam suatu ruang fungsi tertentu sehingga regresi nonparametrik sangat mempertahankan fleksibilitasnya.

Ada beberapa teknik *smoothing* yang digunakan dalam metode nonparametrik antara lain: estimator histogram, kernel, deret orthogonal, spline, k-NN, deret Fourier, dan wavelet (Komang dan Gusti, 2012).

2.2 Estimator Kernel

Estimator kernel merupakan pengembangan dari estimator histogram. Suatu histogram disusun dengan meletakkan titik-titik data ke dalam suatu *bin* atau kelas. Setiap *bin* dinyatakan secara grafik oleh segiempat dengan lebar sama dan tinggi proporsional dengan banyaknya titik-titik data yang terletak dalam *bin* tersebut.

Estimator kernel ini umum digunakan dalam model pendekatan nonparametrik. Hal ini disebabkan estimator densitas mempunyai beberapa kelebihan, yaitu:

- a. Estimator kernel mempunyai bentuk yang fleksibel dan secara matematik mudah dikerjakan.
- b. Estimator kernel mempunyai rata-rata kekonvergenan yang relatif cepat (Hardle, 1990).

Menurut Eubank (1999) pada dasarnya estimator kernel sama dengan estimator linier lainnya hanya saja metode kernel lebih khusus dalam penggunaan metode *bandwidth*.

Bandwidth dinotasikan dengan h adalah konstanta positif untuk menentukan kemulusan dari kurva taksiran regresi. Dengan memilih *bandwidth* yang mendekati 0 maka taksiran yang didapatkan tidak mulus bahkan akan menonjolkan plot datanya saja.

Pada regresi kernel pemilihan *bandwidth* jauh lebih penting daripada pemilihan fungsi kernel. Jika *bandwidth* yang dipilih terlalu kecil maka akan menghasilkan estimasi kurva kurang mulus (*under-smoothing*), sebaliknya jika *bandwidth* terlalu besar maka akan menghasilkan estimasi kurva sangat mulus (*over-smoothing*) yang tidak sesuai dengan pola sebaran data. Sehingga harus dipilih nilai *bandwidth* yang optimum agar dihasilkan estimasi terbaik.

Salah satu teknik regresi nonparametrik yang sering digunakan untuk menaksir fungsi regresi $m(x)$ adalah dengan menggunakan estimator Nadaraya-Watson.

Estimator ini diperoleh dengan menggunakan metode penaksiran fungsi densitas kernel. Bentuk persamaan dari Nadaraya-Watson, yaitu:

$$\hat{m}(x) = \frac{\sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) Y_i}{\sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)} \quad (2.2.1)$$

Keterangan:

$\hat{m}(x)$ = fungsi taksiran regresi

Y_i = data ke- i pada variabel respon Y , dengan $i = 1, 2, 3, \dots, n$

X_i = data ke- i pada variabel prediktor X , dengan $i = 1, 2, 3, \dots, n$

K = fungsi kernel

n = ukuran sampel/banyak pengamatan

h = *bandwidth* atau *smoothing parameter*

2.3 Fungsi Kernel

2.3.1 Definisi Fungsi Kernel

Suatu fungsi kernel harus merupakan fungsi kontinu, berharga riil, simetris, dan terbatas. Menurut Hardle (1994), secara umum fungsi kernel didefinisikan sebagai berikut:

$$K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right) \quad (2.3.1.1)$$

untuk $-\infty < x < \infty, h > 0$

dengan:

K = fungsi kernel

h = *bandwidth* atau *smoothing parameter*

Fungsi kernel tersebut harus memenuhi beberapa syarat, yaitu:

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

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

2.3.2 Macam-Macam Fungsi Kernel

Pada estimator kernel terdapat beberapa fungsi kernel yang umum digunakan untuk estimasi data, seperti pada Tabel 1 berikut:

Tabel 1. Macam-Macam Fungsi Kernel

| No. | Kernel | $K(x)$ |
|-----|---------------------|--|
| 1 | <i>Biweight</i> | $\frac{15}{16}(1-x^2)^2 I(x \leq 1)$ |
| 2 | <i>Triangular</i> | $(1- x)I(x \leq 1)$ |
| 3 | <i>Epanechnikov</i> | $\frac{3}{4}(1-x^2)I(x \leq 1)$ |
| 4 | <i>Rectangular</i> | $\frac{1}{2}I(x \leq 1)$ |

dimana I adalah fungsi indikator, dengan

$$I(x) = \begin{cases} 1 & \text{jika } |x| \leq 1 \\ 0 & \text{jika } |x| > 1 \end{cases}$$

(Sudarno, 2011).

Estimator densitas kernel untuk fungsi densitas $f(x)$ didefinisikan:

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \quad (2.3.2.1)$$

Keterangan:

X_i = data ke- i pada variabel prediktor X , dengan $i = 1, 2, 3, \dots, n$

K = fungsi kernel

n = ukuran sampel/banyak pengamatan

h = *bandwidth* atau *smoothing parameter*

Dari persamaan tersebut dapat dilihat bahwa fungsi \hat{f}_h dipengaruhi oleh fungsi kernel K dan parameter pemulus h . Menurut Indrayanti (2014), parameter pemulus (*bandwidth*) dalam fungsi densitas kernel berfungsi untuk mengatur kehalusan kurva yang akan diestimasi

2.4 Regresi Kernel

Salah satu metode untuk mengestimasi model regresi nonparametrik pada Persamaan (2.1.1) adalah regresi kernel. Regresi kernel merupakan teknik statistik nonparametrik untuk menaksir nilai ekspektasi bersyarat dari suatu variabel acak. Nilai $m(x)$ ekuivalen dengan nilai harapan dari variabel respon jika variabel prediktor $X = x$ telah diketahui, nilai ekspektasi umumnya dinotasikan $E(Y|X = x)$. Regresi kernel bertujuan untuk mendapatkan hubungan *nonlinier* antara variabel X dan Y . Ekspektasi bersyarat Y terhadap X dinyatakan sebagai berikut:

$$E(Y|X = x) = m(X) \text{ atau } \hat{y} = m(x) = \int \frac{yf(x,y)}{f(x)} dy \quad (2.4.1)$$

dimana $f(x, y)$ = fungsi kepadatan bersama dari (X, Y)

$f(x)$ = fungsi kepadatan marginal X

(Suparti, dkk., 2018).

Salah satu teknik regresi nonparametrik yang sering digunakan untuk menaksir fungsi regresi $m(x)$ adalah dengan menggunakan estimator Nadaraya-Watson. Estimator ini diperoleh dengan menggunakan metode penaksiran fungsi densitas kernel. Fungsi densitas peluang bersama $f(x, y)$ diduga dengan perkalian kernel sebagai berikut:

$$\hat{f}(x, y) = \hat{f}_{h_1 h_2}(x, y) = \frac{1}{n} \sum_{i=1}^n K_{h_1}(x - X_i) K_{h_2}(y - Y_i) \quad (2.4.2)$$

Keterangan:

K_{h_1} = fungsi kernel dengan *bandwidth* h_1

K_{h_2} = fungsi kernel dengan *bandwidth* h_2

X_i = data ke- i pada variabel prediktor X untuk kernel K_{h_1} dengan $i = 1, 2, 3, \dots, n$

Y_i = data ke- i pada variabel respon Y untuk kernel K_{h_2} dengan $i = 1, 2, 3, \dots, n$

Taksiran pembilang pada persamaan Nadaraya-Watson dapat diperoleh dari integral fungsi kepadatan bersama $f(x, y)$:

$$\begin{aligned} \int y \hat{f}(x, y) dy &= \frac{1}{n} \sum_{i=1}^n K_{h_1}(x - X_i) \int y K_{h_2}(y - Y_i) dy \\ &= \frac{1}{n} \sum_{i=1}^n K_{h_1}(x - X_i) \int \frac{y}{h_2} K\left(\frac{(y - Y_i)}{h_2}\right) dy \end{aligned}$$

Misal:

$$S = \frac{y - Y_i}{h_2} \quad \text{sehingga} \quad y = Sh_2 + Y_i$$

$$S = \frac{y}{h_2} - \frac{Y_i}{h_2}$$

$$\frac{ds}{dy} = \frac{1}{h_2}$$

$$dy = h_2 ds$$

Kemudian nilai y dan dy disubstitusikan ke dalam persamaan tersebut, sehingga diperoleh

$$\begin{aligned} &= \frac{1}{n} \sum_{i=1}^n K_{h_1}(x - X_i) \int \frac{(Sh_2 + Y_i)}{h_2} K(S) h_2 ds \\ &= \frac{1}{n} \sum_{i=1}^n K_{h_1}(x - X_i) \int (sh_2 + Y_i) K(s) ds \\ &= \frac{1}{n} \sum_{i=1}^n K_{h_1}(x - X_i) \left[\int sh_2 K(s) ds + \int Y_i K(s) ds \right] \\ &= \frac{1}{n} \sum_{i=1}^n K_{h_1}(x - X_i) \left[h_2 \int sK(s) ds + Y_i \int K(s) ds \right] \end{aligned}$$

dimana $\int K(s) ds = 1$

$$= \frac{1}{n} \sum_{i=1}^n K_{h_1}(x - X_i) \left[\left(h_2 \int sK(s) ds \right) + Y_i \right]$$

$\int sK(s) ds + Y_i$ bernilai 0, dimana $K(s)$ merupakan fungsi kernel. Untuk membuktikannya maka disubstitusikan salah satu fungsi kernel, misal *Kernel Rectangular*. Sehingga diperoleh:

$$= \frac{1}{n} \sum_{i=1}^n K_{h_1}(x - X_i) \left[\left(h_2 \int_{-1}^1 \frac{1}{2} s ds \right) + Y_i \right]$$

$$\begin{aligned}
& \text{untuk } (|s| \leq 1) \\
&= \frac{1}{n} \sum_{i=1}^n K_{h_1}(x-X_i) \left[(h_2) \left(\frac{1}{4} s^4 \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right) + Y_i \right] \\
&= \frac{1}{n} \sum_{i=1}^n K_{h_1}(x-X_i) \left[(h_2) \left(\frac{1}{4} (1)^4 - \frac{1}{4} (-1)^4 \right) + Y_i \right] \\
&= \frac{1}{n} \sum_{i=1}^n K_{h_1}(x-X_i) \left[(h_2) \left(\frac{1}{4} - \frac{1}{4} \right) + Y_i \right] \\
&= \frac{1}{n} \sum_{i=1}^n K_{h_1}(x-X_i) [(h_2)(0) + Y_i] \\
&= \frac{1}{n} \sum_{i=1}^n K_{h_1}(x-X_i) Y_i
\end{aligned}$$

Sedangkan taksiran penyebutnya adalah taksiran kepadatan kernel pada Persamaan (2.3.2.1). Sehingga dari kombinasi kedua taksiran probabilitas bersyarat pada Persamaan (2.4.1) akan diperoleh persamaan Nadaraya-Watson, yaitu:

$$\hat{m}(x) = \frac{\frac{1}{n} \sum_{i=1}^n K_h(x-X_i) Y_i}{\frac{1}{n} \sum_{i=1}^n K_h(x-X_i)} \quad (2.4.3)$$

dimana

$$\sum_{i=1}^n K_h(x-X_i) = \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \quad (2.4.4)$$

sehingga

$$\hat{m}(x) = \frac{\sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) Y_i}{\sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)} \quad (2.4.5)$$

dengan K adalah fungsi kernel dan h adalah *bandwidth* atau *smoothing parameter* dan pengontrol kemulusan (Halim dan Bisono, 2006).

Kemudian Persamaan (2.4.5) disubstitusikan dalam Persamaan (2.1.1) sehingga menjadi:

$$Y = \frac{\sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) Y_i}{\sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)} + \varepsilon \quad (2.4.6)$$

2.4.1 Estimator *Kernel Biweight* dengan Metode Nadaraya-Watson

Estimator *Kernel Biweight* didapatkan dengan cara mensubstitusikan rumus fungsi *Kernel Biweight* pada estimator Nadaraya-Watson seperti pada Persamaan (2.4.5) sehingga menjadi:

$$\hat{m}(x) = \frac{\sum_{i=1}^n \left(\frac{15}{16} \left(1 - \left(\frac{x - X_i}{h} \right)^2 \right)^2 I Y_i \right)}{\sum_{i=1}^n \left(\frac{15}{16} \left(1 - \left(\frac{x - X_i}{h} \right)^2 \right)^2 I \right)} \quad (2.4.1.1)$$

Dari Persamaan (2.4.6) akan menjadi:

$$\hat{Y} = \frac{\sum_{i=1}^n \left(\frac{15}{16} \left(1 - \left(\frac{x - X_i}{h} \right)^2 \right)^2 I Y_i \right)}{\sum_{i=1}^n \left(\frac{15}{16} \left(1 - \left(\frac{x - X_i}{h} \right)^2 \right)^2 I \right)} + \varepsilon_i \quad (2.4.1.2)$$

dengan $i = 1, 2, 3, \dots, n$

2.4.2 Estimator *Kernel Triangular* dengan Metode Nadaraya-Watson

Estimator *Kernel Triangular* didapatkan dengan cara mensubstitusikan rumus fungsi *Kernel Triangular* pada estimator Nadaraya-Watson seperti pada Persamaan (2.4.5) sehingga menjadi:

$$\hat{m}(x) = \frac{\sum_{i=1}^n \left(\left(1 - \left| \frac{x - X_i}{h} \right| \right) I Y_i \right)}{\sum_{i=1}^n \left(\left(1 - \left| \frac{x - X_i}{h} \right| \right) I \right)} \quad (2.4.2.1)$$

Dari Persamaan (2.4.6) akan menjadi:

$$\hat{Y} = \frac{\sum_{i=1}^n \left(\left(1 - \left| \frac{x - X_i}{h} \right| \right) I Y_i \right)}{\sum_{i=1}^n \left(\left(1 - \left| \frac{x - X_i}{h} \right| \right) I \right)} + \varepsilon_i \quad (2.4.2.2)$$

dengan $i = 1, 2, 3, \dots, n$

2.4.3 Estimator *Kernel Epanechnikov* dengan Metode Nadaraya-Watson

Estimator *Kernel Epanechnikov* didapatkan dengan cara mensubstitusikan rumus fungsi *Kernel Epanechnikov* pada estimator Nadaraya-Watson seperti pada Persamaan (2.4.5) sehingga menjadi:

$$\hat{m}(x) = \frac{\sum_{i=1}^n \left(\frac{3}{4} \left(1 - \left(\frac{x - X_i}{h} \right)^2 \right) I Y_i \right)}{\sum_{i=1}^n \left(\frac{3}{4} \left(1 - \left(\frac{x - X_i}{h} \right)^2 \right) I \right)} \quad (2.4.3.1)$$

Dari Persamaan (2.4.6) akan menjadi:

$$\hat{Y} = \frac{\sum_{i=1}^n \left(\frac{3}{4} \left(1 - \left(\frac{x - X_i}{h} \right)^2 \right) I Y_i \right)}{\sum_{i=1}^n \left(\frac{3}{4} \left(1 - \left(\frac{x - X_i}{h} \right)^2 \right) I \right)} + \varepsilon_i \quad (2.4.3.2)$$

dengan $i = 1, 2, 3, \dots, n$

2.4.4 Estimator *Kernel Rectangular* dengan Metode Nadaraya-Watson

Estimator *Kernel Rectangular* diperoleh dengan cara mensubstitusikan rumus fungsi *Kernel Rectangular* dalam Estimator Nadaraya-Watson seperti pada Persamaan (2.4.5) sehingga menjadi:

$$\hat{m}(x) = \frac{\sum_{i=1}^n \left(\left(\frac{1}{2} \right) I Y_i \right)}{\sum_{i=1}^n \left(\left(\frac{1}{2} \right) I \right)} \quad (2.4.4.1)$$

dan Persamaan (2.4.6) akan menjadi:

$$\hat{Y} = \frac{\sum_{i=1}^n \left(\left(\frac{1}{2} \right) I Y_i \right)}{\sum_{i=1}^n \left(\left(\frac{1}{2} \right) I \right)} + \varepsilon_i \quad (2.4.4.2)$$

dengan $i = 1, 2, 3, \dots, n$

2.5 Pemilihan *Bandwidth* Optimum

Bandwidth dinotasikan dengan h adalah konstanta positif untuk menentukan kemulusan dari kurva taksiran regresi. Dengan memilih *bandwidth* yang mendekati 0 maka taksiran yang didapatkan tidak mulus bahkan akan menonjolkan plot datanya saja.

Pada regresi kernel pemilihan *bandwidth* jauh lebih penting daripada pemilihan fungsi kernel. Jika *bandwidth* yang dipilih terlalu kecil maka akan menghasilkan estimasi kurva kurang mulus (*under-smoothing*), sebaliknya jika *bandwidth* terlalu besar maka akan menghasilkan estimasi kurva sangat mulus (*over-smoothing*) yang tidak sesuai dengan pola sebaran data. Sehingga harus dipilih nilai *bandwidth* yang optimum agar dihasilkan estimasi terbaik.

Terdapat beberapa metode yang digunakan dalam pemilihan *bandwidth* optimum, salah satunya adalah menggunakan kriteria *Generalized Cross Validation* (GCV) (Galub, dkk., 1979), didefinisikan dengan:

$$GCV = \frac{MSE}{\left(\frac{1}{n} \text{tr}(\mathbf{I} - \mathbf{H}(\mathbf{h}))\right)^2} \quad (2.5.1)$$

dengan

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{m}(x_i))^2$$

$$\mathbf{H}(\mathbf{h}) = \mathbf{X}(\mathbf{X}'\mathbf{X} + n\mathbf{h}\mathbf{I})^{-1}\mathbf{X}'$$

$\hat{m}(x_i)$ = fungsi taksiran regresi pada data taksiran ke- i

Y_i = data ke- i pada variabel respon Y , dengan $i = 1, 2, 3, \dots, n$

n = banyak data

h = *bandwidth*

\mathbf{tr} = penjumlahan elemen dari diagonal utama suatu matriks berordo $n \times n$

\mathbf{X} = matriks data

\mathbf{I} = matriks identitas

Menurut Komang dan Gusti (2012), kebaikan suatu estimator dapat dilihat dari tingkat kesalahannya. Terdapat beberapa kriteria untuk menentukan estimator terbaik dalam model regresi nonparametrik, diantaranya:

a. *Mean Square Error (MSE)*

Untuk mengukur *error* biasanya digunakan *Mean Square Error*. Estimator terbaik dipilih berdasarkan nilai MSE terkecil. *Mean Square Error (MSE)* adalah rata-rata dari kuadrat kesalahan.

$$MSE = \frac{1}{n} \sum_{i=1}^n e_i^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{m}(x_i))^2 \quad (2.5.2)$$

b. *Root Mean Square Error (RMSE)*

$$RMSE = \sqrt{MSE} \quad (2.5.3)$$

c. *Mean Absolute Deviation (MAD)*

$$MAD = \frac{1}{n} \sum_{i=1}^n |e_i| = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| \quad (2.5.4)$$

keterangan:

$\hat{m}(x_i)$ = fungsi taksiran regresi pada data taksiran ke- i

n = banyak data

Y_i = data ke- i pada variabel respon Y , dengan $i = 1, 2, 3, \dots, n$

\hat{Y}_i = nilai prediksi dari variabel Y_i

2.6 Pajak Kendaraan Bermotor

Pajak adalah pungutan terhadap masyarakat oleh negara berdasarkan undang-undang yang bersifat memaksa dan terutang yang wajib dibayar dengan tidak mendapat imbalan secara langsung, yang hasilnya digunakan untuk membiayai pengeluaran-pengeluaran negara dalam penyelenggaraan pemerintahan dan pembangunan (Siahaan, 2013).

Pajak kendaraan bermotor menurut Undang-Undang No. 28 Tahun 2009 tentang pajak daerah dan retribusi daerah adalah “Pajak Kendaraan Bermotor adalah pajak atas kepemilikan dan/atau penguasaan kendaraan bermotor”. Berdasarkan Undang-Undang tersebut pasal 5 mengenai dasar pengenaan pajak kendaraan bermotor adalah hasil perkalian dari dua unsur pokok, yaitu nilai jual kendaraan bermotor dan bobot yang mencerminkan kadar kerusakan jalan dan pencemaran lingkungan akibat penggunaan kendaraan bermotor tersebut. Nilai jual kendaraan bermotor sesuai dengan harga pasar kendaraan bermotor, jenis kendaraan bermotor, merk kendaraan bermotor, tahun pembuatan kendaraan bermotor, berat total kendaraan bermotor, serta dokumen impor jenis kendaraan tertentu.

Subjek Pajak Kendaraan Bermotor adalah orang pribadi, Badan Pemerintah, Pemerintah Daerah, TNI, dan Polri yang memiliki dan/atau menguasai Kendaraan Bermotor. Sementara itu wajib Pajak Kendaraan Bermotor adalah orang pribadi, Badan Pemerintah, Pemerintah Daerah, TNI, dan Polri yang memiliki dan/atau menguasai Kendaraan Bermotor dan/atau kendaraan khusus atau alat-alat berat dan besar.

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilakukan pada semester ganjil tahun akademik 2019/2020 di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

3.2 Data Penelitian

Data yang digunakan pada penelitian ini adalah data sekunder yang bersumber dari Kantor Badan Pendapatan Daerah (BAPENDA) Provinsi Lampung yaitu jumlah Penerimaan Pajak Kendaraan Bermotor (PKB) per bulan yang dihimpun dari bulan Januari 2015 sampai Desember 2017 dapat dilihat pada Tabel 2 berikut:

Tabel 2. Data Penerimaan Pajak Kendaraan Bermotor

| Bulan | PKB (Rupiah) | | |
|-----------|----------------|----------------|----------------|
| | 2015 | 2016 | 2017 |
| Januari | 45.766.583.075 | 46.446.495.475 | 53.451.232.614 |
| Februari | 40.530.015.327 | 42.856.251.120 | 44.517.868.057 |
| Maret | 47.134.078.083 | 46.720.850.203 | 53.621.494.247 |
| April | 46.510.074.322 | 46.416.443.704 | 47.496.930.590 |
| Mei | 41.782.051.060 | 48.229.788.345 | 51.718.722.693 |
| Juni | 50.579.743.166 | 55.508.342.045 | 45.935.243.298 |
| Juli | 48.102.168.986 | 40.631.954.210 | 58.435.088.593 |
| Agustus | 53.831.500.159 | 56.283.725.672 | 60.267.961.807 |
| September | 53.793.159.122 | 54.338.391.200 | 52.684.460.273 |
| Oktober | 51.045.512.048 | 48.498.087.989 | 62.587.354.915 |
| November | 47.004.000.833 | 49.356.670.241 | 68.590.047.345 |
| Desember | 44.389.753.073 | 48.926.376.581 | 77.726.219.634 |

3.3 Metode Penelitian

Langkah-langkah dalam mengestimasi data penerimaan pajak kendaraan bermotor untuk mendapatkan kurva estimasi menggunakan estimator *Biweight*, *Triangular*, *Epanechnikov*, dan *Rectangular* sebagai berikut:

1. Pengambilan data penerimaan pajak kendaraan bermotor pada Kantor Badan Pendapatan Daerah Provinsi Lampung.
2. Menerapkan regresi kernel pada data penelitian yang telah diperoleh dengan membuat metode regresi *Kernel Biweight*, *Triangular*, *Epanechnikov*, dan *Rectangular* menggunakan *software RStudio Version 1.1.419*.

Algoritma yang digunakan dalam membuat program kernel pada *software RStudio Version 1.1.419* sebagai berikut:

- a. Data yang telah diimport ke program *RStudio Version 1.1.419* akan tersedia sebagai data x_i , x_j , dan y_j , dimana $i = 1, 2, 3, \dots, n$ dan $j = 1, 2, 3, \dots, n$ dengan: $x_i = x$ (data ke- i), $x_j = X_i$ (data ke- i pada variabel prediktor X), dan $y_j = Y_i$ (data ke- i pada variabel respon Y).
- b. Mendefinisikan jumlah data yang akan diteliti sebanyak 36 data dan *bandwidth* yang akan dimasukkan dalam program. Dalam mendefinisikan nilai *bandwidth* akan dimulai dari nilai *bandwidth* terkecil yaitu 0,1 sampai *bandwidth* sebesar 10 dengan menggunakan selang kenaikan sebesar 0,1. Pengaplikasian batas *bandwidth* ini dilakukan secara *trial and error*, sehingga diharapkan nilai *bandwidth* antara 0,1 sampai 10 dengan menggunakan selang kenaikan sebesar 0,1 dapat ditemukan nilai *bandwidth* yang paling optimum.

c. Setelah semua terdefinisi, hitung kernelnya dengan tahapan berikut:

1) Untuk $i = 1$

a) Terlebih dahulu akan dihitung $\left(\frac{x_i - x_j}{h}\right)$ dengan $j = 1, 2, 3, \dots, n$

sehingga menjadi $\left(\frac{x_1 - x_1}{h}\right), \left(\frac{x_1 - x_2}{h}\right), \left(\frac{x_1 - x_3}{h}\right), \dots, \left(\frac{x_1 - x_n}{h}\right)$.

b) Selanjutnya dihitung $K\left(\frac{x_1 - x_j}{h}\right)$ dengan K adalah jika

(1) Fungsi *Kernel Biweight* maka

$$K\left(\frac{x_1 - x_j}{h}\right) = \frac{15}{16} \left(1 - \left(\frac{x_1 - x_j}{h}\right)^2\right)^2 I$$

(2) Fungsi *Kernel Triangular* maka

$$K\left(\frac{x_1 - x_j}{h}\right) = \left(1 - \left|\frac{x_1 - x_j}{h}\right|\right) I$$

(3) Fungsi *Kernel Epanechnikov* maka

$$K\left(\frac{x_1 - x_j}{h}\right) = \frac{3}{4} \left(1 - \left(\frac{x_1 - x_j}{h}\right)^2\right) I$$

(4) Fungsi *Kernel Rectangular* maka

$$K\left(\frac{x_1 - x_j}{h}\right) = \left(\frac{1}{2}\right) I$$

dengan syarat jika $\left|\frac{x_1 - x_j}{h}\right| \leq 1$ maka $I = 1$ dan jika $\left|\frac{x_1 - x_j}{h}\right| > 1$

maka $I = 0$.

(5) Dihitung pula $K\left(\frac{x_1 - x_j}{h}\right) y_j$ dengan $j = 1, 2, 3, \dots, n$.

(6) Kemudian dihitung

$$\hat{m}(x) = \frac{\sum_{i=1}^n K\left(\frac{x_1 - x_j}{h}\right) y_j}{\sum_{i=1}^n K\left(\frac{x_1 - x_j}{h}\right)}$$

Untuk

(1) Fungsi *Kernel Biweight* maka

$$K\left(\frac{x_1 - x_j}{h}\right) = \frac{15}{16} \left(1 - \left(\frac{x_1 - x_j}{h}\right)^2\right)^2 I$$

(2) Fungsi *Kernel Triangular* maka

$$K\left(\frac{x_1 - x_j}{h}\right) = \left(1 - \left|\frac{x_1 - x_j}{h}\right|\right) I$$

(3) Fungsi *Kernel Epanechnikov*

$$K\left(\frac{x_1 - x_j}{h}\right) = \frac{3}{4} \left(1 - \left(\frac{x_1 - x_j}{h}\right)^2\right) I$$

(4) Fungsi *Kernel Rectangular*

$$K\left(\frac{x_1 - x_j}{h}\right) = \left(\frac{1}{2}\right) I$$

2) Selanjutnya untuk $i = 2, 3, \dots, n$ tahapannya sama dengan $i = 1$

3. Menentukan *bandwidth* yang digunakan dalam estimator.
4. Memasukkan *bandwidth* ke dalam estimator *Kernel Biweight*, *Triangular*, *Epanechnikov*, dan *Rectangular*.
5. Mengetahui nilai GCV dari *bandwidth* yang telah digunakan dalam estimator kernel.
6. Memilih *bandwidth* optimum berdasarkan nilai GCV minimum dengan rumus sebagai berikut:

$$GCV = \frac{MSE}{\left(\frac{1}{n} \text{tr}(I - H(h))\right)^2} \quad (2.5.1)$$

dengan

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{m}(x_i))^2$$

$$H(\mathbf{h}) = X(X'X + n\mathbf{h}I)^{-1}X'$$

$\hat{m}(x_i)$ = fungsi taksiran regresi pada data taksiran ke- i

Y_i = data ke- i pada variabel respon Y , dengan $i = 1, 2, 3, \dots, n$

n = banyak data

h = *bandwidth*

\mathbf{tr} = penjumlahan elemen dari diagonal utama suatu matriks berordo
 $n \times n$

X = matriks data

I = matriks identitas

7. Membandingkan hasil estimasi antara estimator *Kernel Biweight*, *Triangular*, *Epanechnikov*, dan *Rectangular* menggunakan *bandwidth* yang optimum.
8. Menentukan estimator terbaik bagi data sekunder menggunakan *bandwidth* yang optimum berdasarkan kriteria nilai GCV minimum dengan nilai MSE terkecil.

V. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan pembahasan pada Bab 4 dapat disimpulkan bahwa dari keempat estimator kernel yang diperoleh, nilai MSE yang paling kecil berada pada estimator *Kernel Triangular* dengan nilai $1,536755 \times 10^{19}$ untuk *bandwidth* optimum sebesar 4,8. Sehingga hasil dari perbandingan nilai MSE dari keempat estimator kernel tersebut dapat disimpulkan bahwa estimator *Kernel Triangular* merupakan estimator terbaik untuk data penerimaan pajak kendaraan bermotor dibandingkan dengan estimator *Kernel Biweight*, *Epanechnikov*, dan *Rectangular* karena memiliki tingkat kesalahan yang paling kecil

5.1 Saran

Berdasarkan analisis data dan pembahasan, saran yang dapat diberikan untuk penelitian selanjutnya sebagai berikut:

1. Menggunakan pendekatan model regresi nonparametrik lain untuk mendekati pola sebaran data penerimaan pajak kendaraan bermotor seperti regresi spline, deret Fourier, wavelet atau lainnya.
2. Menggunakan estimator lain, seperti Priestley-Chao dan Gasser-Muller.

DAFTAR PUSTAKA

- Eubank, R. 1999. *Spline Smoothing and Nonparametrik Regression*. Marcel Dekker, New York.
- Galub, G.H., Heath, M., & Wahba, G. 1979. Generalized Cross-Validation as A Method for Choosing a Good Ridge Parameter. *Technometrics*. **21**(2): 215-223.
- Halim, S. & Bisono, I. 2006. Fungsi-Fungsi Kernel pada Metode Regresi Nonparametrik dan Aplikasinya pada Priest River Experimental Forest's Data. *Jurnal Teknik Industri*. **8**(1): 73-81.
- Hardle, W. 1990. *Smoothing Techniques with Implementation in S*. Springer Verlag, New York.
- Hardle, W. 1994. *Applied Nonparametric Regression*. Cambridge University Press, Cambridge.
- Indrayanti, A.I. 2014. Estimator Kernel Cosinus dan Kernel Gaussian dalam Model Regresi Nonparametrik pada Data Butterfly Diagram Siklus Aktivitas ke-23 (Studi Kasus di BPD LAPAN Watukosek). Skripsi. UIN Maulana Malik Ibrahim Malang, Malang.
- Komang, G. & Gusti, A. 2012. Estimator Kernel dalam Model Regresi Nonparametrik. *Jurnal Matematika*. **2**(1): 19-30.
- Siahaan. 2013. *Pajak Daerah dan retribusi Daerah*. Raja Grafindo Persada, Jakarta.

Sudarno. 2011. Pemulusan Sebaran Data Menggunakan Penaksiran Kernel Nadaraya-Watson dan Linier Lokal untuk Kernel Normal. *Prosiding Seminar Nasional Statistik*, 2011. Universitas Diponegoro.

Suparti, Rukun, S., Alan, P., & Alvita, R.D. 2018. *Regresi Nonparametrik*. Wade Group, Jawa Timur.