PENENTUAN PARAMETER PENGHALUS *SMOOTHING SPLINE* DALAM REGRESI SEMIPARAMETRIK DENGAN GCV

SKRIPSI

Oleh:
LUSIANA WATI
NIM. 09610101

JURUSAN MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2014

PENENTUAN PARAMETER PENGHALUS *SMOOTHING SPLINE* DALAM REGRESI SEMIPARAMETRIK DENGAN GCV

SKRIPSI

Diajukan kepada:
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
untuk Memenuhi Salah Satu Persyaratan
dalam Memperoleh Gelar Sarjana Sains (S.Si)

Oleh:
LUSIANA WATI
NIM. 09610101

JURUSAN MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG
2014

PENENTUAN PARAMETER PENGHALUS *SMOOTHING SPLINE* DALAM REGRESI SEMIPARAMETRIK DENGAN GCV

SKRIPSI

Oleh:
LUSIANA WATI
NIM. 09610101

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 7 April 2014

Pembimbing I,	Pembimbing II,
Dr. Sri Harini, M.Si	Ari Kusumastuti, S.Si, M.Pd
NIP. 19731014 200112 2 002	NIP. 19770521 200501 2 004

Mengetahui,
Ketua Jurusan Matematika

Dr. Abdussakir, M.Pd
NIP. 19751006 200312 1 001

PENENTUAN PARAMETER PENGHALUS *SMOOTHING SPLINE* DALAM REGRESI SEMIPARAMETRIK DENGAN GCV

SKRIPSI

oleh:
LUSIANA WATI
NIM. 09610101

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima sebagai Salah Satu Persyaratan
untuk Memperoleh Gelar Sarjana Sains (S.Si)

Tanggal: 11 April 2014

Susunan Dewan Penguji		**Tanda Tangan**
1. Penguji Utama	: Drs. H. Turmudi, M.Si NIP. 19571005 198203 1 006	________________
2. Ketua Penguji	: Evawati Alisah, M.Pd NIP. 19720604 199903 2 001	________________
3. Sekretaris Penguji	: Dr. Sri Harini, M.Si NIP. 19731014 200112 2 002	________________
4. Anggota Penguji	: Ari Kusumastuti, S.Si, M.Pd NIP. 19770521 200501 2 004	________________

Mengetahui dan Mengesahkan,
Ketua Jurusan Matematika,

Dr. Abdussakir, M.Pd
NIP. 19751006 200312 1 001

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : LUSIANA WATI

NIM : 09610101

Jurusan : Matematika

Fakultas : Sains dan Teknologi

Judul : Penentuan Parameter Penghalus *Smoothing Spline* dalam Regresi Semiparametrik dengan GCV

menyatakan dengan sebenarnya bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambilalihan data, tulisan atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka. Apabila di kemudian hari terbukti atau dapat dibuktikan skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 7 April 2014

Yang membuat pernyataan,

Lusiana Wati

NIM. 09610101

MOTTO

إِلا مَنْ تَابَ وَآمَنَ وَعَمِلَ صَالِحًا فَأُولَئِكَ يَدْخُلُونَ الْجَنَّةَ وَلا يُظْلَمُونَ شَيْئًا (٦٠)
جَنَّاتِ عَدْنٍ الَّتِي وَعَدَ الرَّحْمَنُ عِبَادَهُ بِالْغَيْبِ إِنَّهُ كَانَ وَعْدُهُ مَأْتِيًّا (٦١)

(60) *kecuali orang yang bertaubat, beriman dan beramal saleh, maka mereka itu akan masuk surga dan tidak dianiaya (dirugikan) sedikitpun,*

(61) *yaitu surga 'Adn yang telah dijanjikan oleh Tuhan Yang Maha Pemurah kepada hamba-hamba-Nya, sekalipun (surga itu) tidak nampak. Sesungguhnya janji Allah itu pasti akan ditepati.*

(Surat Maryam:60-61)

PERSEMBAHAN

بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِيْمِ

Alhamdulillahi Robbil 'alamin

Tiada Kata yang Pantas untuk diucapkan selain bersyukur atas nikmat dan karunia Allah SWT, maka penulis persembahkan Skripsi ini kepada:

Bapak dan Ibu Tercinta

Bapak Abdul Rahman dan Ibu Sunarsih

Kakak Tercinta

Zuris Diana Susanti dan Zaki Bahrul Rifai

Keponakan Tercinta

Vera Putri Anggraeni

Sahabat Terbaik

Amalia Intifaada, Mahfud Tono Tiyanto dan Tegar Isnain Sulistyo

KATA PENGANTAR

Assalamu'alaikum Wr.Wb.

Syukur *Alhamdulillah* penulis haturkan ke hadirat Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan studi di Jurusan Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Selanjutnya penulis haturkan ucapan terima kasih seiring do'a dan harapan *jazakumullah ahsanal jaza'* kepada semua pihak yang telah membantu menyelesaikan skripsi ini. Ucapan terima kasih ini penulis sampaikan kepada:

1. Prof. Dr. H. Mudjia Rahardjo, M.Si, selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang yang telah banyak memberikan pengetahuan dan pengalaman yang berharga.
2. Dr. drh. Hj. Bayyinatul Muchtaromah, M.Si, selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Abdussakir, M.Pd, selaku Ketua Jurusan Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Dr. Sri Harini, M.Si dan Ari Kusumastuti, S.Si, M.Pd, selaku dosen pembimbing skripsi, yang telah meluangkan waktunya untuk memberikan arahan dan bimbingan selama penulisan skripsi ini.
5. Segenap sivitas akademika Jurusan Matematika, terutama seluruh dosen, terima kasih atas segenap ilmu dan bimbingannya.

6. Kedua orang tua, yang tak henti-hentinya memanjatkan do’a serta bekerja memeras keringat untuk pendidikan, kebahagiaan, dan kesuksesan masa depan penulis.
7. Kakak Zuris Diana Susanti dan Zaki Bahrul Rifai, serta keponakan tercinta yang telah memberikan semangat dan dukungan kepada penulis.
8. Kakak Mahfud Tono Tiyanto dan Tegar Isnain Sulistyo yang telah memberikan motivasi kepada penulis.
9. Sahabat-sahabat senasib seperjuangan mahasiswa Jurusan Matematika angkatan 2009, khususnya Amalia Intifaada, Rohatul Warda, Farida Ulin Nuha, Amanatul Husnia, Misbah, Lala, Mas Munawir, Mbak Ratna, Mbak Mifta, dan Anis Safidah terima kasih atas segala pengalaman berharga dan kenangan terindah saat menuntut ilmu bersama.
10. Semua pihak yang tidak mungkin penulis sebut satu persatu, terima kasih atas keikhlasan bantuan moril dan spiritual yang sudah diberikan kepada penulis.

Penulis berharap semoga skripsi ini dapat memberikan manfaat kepada para pembaca khususnya bagi penulis secara pribadi. *Amin Ya Rabbal Alamin.*

Wassalamu’alaikum Wr.Wb.

Malang, April 2014

Penulis

DAFTAR ISI

HALAMAN JUDUL
HALAMAN PENGAJUAN
HALAMAN PERSETUJUAN
HALAMAN PENGESAHAN
HALAMAN PERNYATAAN KEASLIAN TULISAN
HALAMAN MOTTO
HALAMAN PERSEMBAHAN
KATA PENGANTAR viii
DAFTAR ISI x
DAFTAR GAMBAR xii
DAFTAR TABEL xiii
DAFTAR LAMPIRAN xiv
ABSTRAK xv
ABSTRACT xvi
ملخص البحث xvii

BAB I PENDAHULUAN
1.1 Latar Belakang 1
1.2 Rumusan Masalah 8
1.3 Tujuan Penelitian 8
1.4 Batasan Masalah 9
1.5 Manfaat Penelitian 9
1.6 Metode Penelitian 10
1.6.1 Pendekatan Penelitian 10
1.6.2 Langkah-langkah Penelitian 10
1.7 Sistematika Penulisan 11

BAB II TINJAUAN PUSTAKA
2.1 Regresi Parametrik 13
2.2 Regresi Nonparametrik 13
2.3 Regresi Semiparametrik 14
2.4 Estimasi Parameter 15
2.4.1 Metode *Least Square* 16
2.4.2 Metode Newton Raphson 17
2.5 Asumsi-asumsi Analisis Regresi Semiparametrik 21
2.6 Uji Asumsi pada *Error* 21
2.6.1 Uji Asumsi Kenormalan 21
2.6.2 Uji Asumsi Non-multikolinearitas 22
2.6.3 Uji Asumsi Non-autokorelasi 23
2.6.4 Uji Asumsi Kehomogenan Ragam *Error* 24
2.7 *P-Value* 25
2.8 Kelayakan Model 26

2.9 *Spline* 26
2.10 *Spline Cubic* 27
2.11 Mencari Titik Knot 28
2.12 Model *Smoothing Spline* dalam Regresi Semiparametrik 29
2.13 Estimator Fungsi *Smoothing Spline* dalam Regresi Semiparametrik 31
2.14 Pemilihan Parameter Penghalus λ Optimal 33
2.15 Menguji Signifikasi Model Regresi Semiparametrik 35
2.15.1 Uji Simultan (Uji *F*) 35
2.15.2 Uji Individu atau Uji Parsial (Uji *t*) 36
2.16 Kemiskinan 36
2.17 Tingkat Pengangguran Terbuka 38
2.18 Indeks Pendidikan 39
2.19 Inspirasi Al-Qur'an dalam Kajian tentang Estimasi 40

BAB III PEMBAHASAN
3.1 Optimasi *Penalized Least Square (PLS)* 42
3.2 Estimasi Parameter β 42
3.3 Estimasi Parameter f 43
3.4 Menentukan $\hat{\beta}_\lambda$ dan $\hat{f}_\lambda$ 43
3.5 Menentukan *Hat Matrix* $A(\lambda)$ 45
3.6 Menentukan Parameter Penghalus Optimal dengan Metode GCV 45
3.7 Regrsi Semiparametrik untuk Memodelkan Kemiskinan di Kota Malang pada Tahun 2005-2012 46
3.7.1 Deskripsi Data 46
3.7.2 Mendeteksi Variabel Prediktor Komponen Parametrik dan Variabel Nonparametrik Menggunakan *Scatter Plot* 47
3.7.3 Memilih Model Tentatif untuk *Smoothing Spline Cubic* 49
3.8 Uji Signifikansi Model Regresi Semiparametrik 51
3.8.1 Uji Simultan 51
3.8.2 Uji Asumsi 51
3.8.3 Uji Kelayakan Model 52
3.9 Inspirasi Al-Qur'an dalam Kajian tentang Estimasi Parameter 53

BAB IV PENUTUP
4.1 Kesimpulan 55
4.2 Saran 56

DAFTAR PUSTAKA 57

LAMPIRAN

DAFTAR GAMBAR

Gambar 2.1 Grafik Newton Raphson 17
Gambar 2.2 Grafik Newton Raphson Fungsi Nonparametrik 20
Gambar 3.1 *Plot* untuk Uji Normalitas pada Tingkat Pengangguran Terbuka.. 47
Gambar 3.2 *Plot* untuk Uji Normalitas pada Indeks Pendidikan 48

DAFTAR TABEL

Tabel 2.1 Ketentuan Durbin Watson 24
Tabel 3.1 Tabel 3.1 Model Tentatif *Smoothing Spline Cubic* dengan semua kemungkinan satu parameter penghalus dan nilai *MSE* beserta *GCV*nya 49
Tabel 3.2 ANOVA Regresi *Smoothing Spline Cubic* Satu Parameter Penghalus 51

DAFTAR LAMPIRAN

Lampiran 1 Data Pengamatan
Lampiran 2 Hasil Perhitungan *MSE* dengan MATLAB
Lampiran 3 Hasil Perhitungan *GCV* dengan MATLAB
Lampiran 4 Hasil Perhitungan *Hat Matrix* dengan MATLAB
Lampiran 5 *Output* MINITAB

ABSTRAK

Wati, Lusiana. 2014. **Penentuan Parameter Penghalus *Smoothing Spline* pada Regresi Semiparametrik dengan GCV**. Skripsi. Jurusan Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
Pembimbing: (I) Dr. Sri Harini, M.Si
(II) Ari Kusumastuti, S.Si, M.Pd

Kata kunci: Parameter Penghalus, *Smoothing Spline*, *GCV*, Semiparametrik, *Penalized Least Square (PLS).*

Metode GCV digunakan untuk menentukan parameter penghalus yang optimal pada regresi semiparametrik dengan pendekatan *smoothing spline* yang berhubungan dengan nilai *MSE*, pemilihan titik knot dan *hat matrix*. Pemilihan titik knot pada hakekatnya memilih parameter penghalus. Ketika nilai parameter penghalus λ kecil, maka akan terbentuk kurva regresi yang kasar dan untuk λ yang semakin besar maka kurva regresinya akan semakin halus. Nilai λ adalah $0 < \lambda < 1$ yang ditentukan secara random dengan bantuan program MATLAB. Untuk mendapatkan nilai GCV, maka perlu adanya estimasi parameter untuk membentuk *hat matrix* yang digunakan untuk menghitung nilai GCV. *Hat matrix* dibentuk dari $\hat{\beta}$ dan $\hat{f}$ yang disubstitusikan ke dalam regresi semiparametrik. Estimasi parameter untuk model regresi semiparametrik diperoleh dengan optimasi *PLS* (*Penalized Least Square*). Ketika nilai GCV yang optimal sudah didapatkan, maka akan terbentuk model regresi hasil estimasi yang terbaik. Parameter penghalus optimal dengan cara memilih λ yang memiliki nilai GCV paling minimum. Dari hasil penelitian, nilai GCV paling minimum bernilai 1.4350 pada saat $\lambda = 0.99$. Sehingga model tentatif regresi semiparametrik *smoothing spline cubic* dengan satu parameter penghalus untuk jumlah kemiskinan di Kota Malang yang dipengaruhi oleh tingkat pengangguran terbuka dan indeks pendidikan adalah

$$\hat{Y} = 5.8654 + 0.0697X + 0.0329 - 0.8395t + 0.5958t^2 + 2.3469t^3 + 0.6524(t - 0.99)^3.$$

Dari beberapa uji untuk model tentatif tersebut disimpulkan bahwa adanya autokorelasi antar *error* dan terjadi non-multikolinieritas.

ABSTRACT

Wati, Lusiana. 2014. **The Determination of Smoothing Spline Smoothing Parameter in semiparametric regression with GCV**. Thesis. Department of Mathematic The Faculty of Science and Technology Maulana Malik Ibrahim State Islamic University Malang.
Advisors: (I) Dr. Sri Harini, M.Si
(II) Ari Kusumastuti, S.Si, M.Pd

Keywords: Smoothing Parameter, Smoothing Spline, GCV, semiparametric, Penalized Least Square (*PLS*).

GCV method is used to determine the optimal smoothing parameter in semiparametric regression with smoothing spline approach which is related with the *MSE* (Mean Square Error), the selection of knots point and hat matrix. In fact, the knot points selection is dealing with choosing the smoothing parameter. When the smoothing parameter value λ is small, it will form the rough regression curve and for greater λ, the regression curve will be smoother. The value of λ $(0<\lambda<1)$ is determined randomly by MATLAB. To obtain the GCV value, it is necessary to estimate the parameters for constructing the hat matrix which is used to calculate the value of GCV. Hat matrix is obtained from $\hat{\beta}$ and $\hat{f}$ that are substituted to semiparametric regression. Parameter estimation for semiparametric regression models is obtained from *PLS* (Penalized Least Square) optimization. When the optimal value of GCV has been established, it will obtain the regression model form the best regression model of the best results. The optimum smoothing parameter is obtained by choosing the λ that has minimum GCV value. From the research, we got the minimum GCV value worth 1.4350 at the time $\lambda = 0.99$. So the tentative model of semiparametric regression of smoothing spline cubic with one smoothing parameter for the amount of poverty in the city of Malang are influenced by the unemployment rate and education index is
$\hat{Y} = 5.8654 + 0.0697X + 0.0329 - 0.8395t + 0.5958t^2 + 2.3469t^3 + 0.6524(t-0.99)^3$.
From the tests of tentative model, we can conclude that there is the presence of autocorrelation between errors and the non-multicollinearity.

ملخص البحث

واتي ، لوسييانا. عام٢٠١٤. التنعيم المفتاح التنعيم المعلمة تحديد الانحدار نصف على متري مع . GCV البحث العلمي من قسم الرياضيات كلية العلوم وتكنولوجيا مولانا مالك إبراهيم مالانج الدولة الإسلامية.

المشرف : (١) دكتور أندوس. سري هاريني الماجستير

(٢) آري كوسوماستوتي ، الماجستير

الكلمات الرئيسية : معلمة التنعيم ، التنعيم المفتاح ، GCV، شبه حدودي ، معاقب ساحة الأقل. (PLS) يستخدم طر يقة GCV لتحديد المعلمة تجانس الأمثل في شبه الانحدار حدودي مع تجانس النهج سين الذي يرتبط مع MSE(متوسط مربع الخطأ) ، واختيار نقطة عقدة، و قبعة المصفوفة .في الواقع ، اختيار نقاط عقدة يتعامل مع اختيار المعلمة المتجا نس . عندما تكون قيمة معلمة تجانس صغير ، فإنه مشكون تشكيل منحنى الانحدار وءر و وإنكانت قيحترمملمترتجابسهاكبرةفتكونتشكيلهاسلستر . قيمتر λ $(0 < \lambda < 1)$ تحددءشوائياباستخدام MATLAB. للحصول على قيمة GCV ، فمن الضروري لتقدير المعلمات لبناء مصفوفة قبعة الذي يستخدم لحساب قيمةGCV .يتم الحصول على قبعةمصفوفةمن $\hat{\beta}$ و $\hat{f}$ التي يتم استبداله لشبه حدودي التراجع.يتم الحصول على تقدير المعلمة ل نماذج شبه حدودالتراجع من أمثلPLS (معاقب بأقل ساحة) .عندما تم تأسيس قيمة مثلى من GCV ، فيتم الحصول تشكيل نموذج التراجع أفضل نموذج التراجع من أفضل النتائج .يتم الحصول على المعلمة تجانس الأمثل عن طريق اختيار لديه الحد الأدنى من قيمة GCV. من بحث ، حصلنا على قيمة الحد الأدنى GCV بقيمة 1.4350 في الوقت .لذلك تتأثر نموذج مبدئي شبه حد ود التراجع من تجانس الشريحة مكعب واحد مع المعلمة تجانس لمبلغ من الفقر في مدينة مالانج تلاه مؤشر معدل البطالة و التعليم:

$$\hat{Y} = 5.8654 + 0.0697X + 0.0329 - 0.8395t + 0.5958t^2 + 2.3469t^3 + 0.6524(t - 0.99)^3$$

من الاختبارات لنموذج مبدئي ، يمكننا أن نستنتج أن هناك وجود الارتباط بين الأخطاء وعدم الخطية المتعددة.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam kehidupan sehari-hari tidak terlepas dari statistik yang menyangkut tentang data, fakta dan informasi. Salah satu kegiatan dalam statistik adalah pengumpulan data seperti yang diisyaratkan dalam surat Yunus (10) ayat 61 yang membicarakan mengenai pengumpulan data yang berbunyi:

وَمَا تَكُونُ فِي شَأْنٍ وَمَا تَتْلُو مِنْهُ مِنْ قُرْآنٍ وَلَا تَعْمَلُونَ مِنْ عَمَلٍ إِلَّا كُنَّا عَلَيْكُمْ شُهُودًا إِذْ تُفِيضُونَ فِيهِ ۚ وَمَا يَعْزُبُ
عَنْ رَبِّكَ مِنْ مِثْقَالِ ذَرَّةٍ فِي الْأَرْضِ وَلَا فِي السَّمَاءِ وَلَا أَصْغَرَ مِنْ ذَٰلِكَ وَلَا أَكْبَرَ إِلَّا فِي كِتَابٍ مُبِينٍ

Artinya:
"Kamu tidak berada dalam suatu keadaan dan tidak membaca suatu ayat dari Al-Qur'an dan kamu tidak mengerjakan suatu pekerjaan, melainkan kami menjadi saksi atasmu di waktu kamu melakukannya. Tidak luput dari pengetahuan Tuhanmu biarpun sebesar zarah (atom) di bumi ataupun di langit. Tidak ada yang lebih kecil dan tidak pula yang lebih besar dari pada itu, melainkan (semua tercatat) dalam kitab yang nyata (Lawh Mahfuz)" (QS. Yunus: 61).

Dalam ayat tersebut dijelaskan bahwa pada saat manusia melaksanakan amal perbuatannya tidak ada yang terlepas dari pengawasan Allah. Dia menyaksikan semua amal perbuatan itu pada saat melakukannya meskipun amalan tersebut lebih kecil dari benda yang terkecil, bahkan urusan itu maha penting sehingga tak terkendalikan oleh manusia. Ilmu Allah tidak hanya meliputi segala macam urusan yang ada di bumi yang kebiasaannya urusan ini dapat dibayangkan oleh mereka secara mudah. Juga meliputi segala macam urusan di langit yang urusannya lebih sulit tergambar dalam pikiran mereka. Di akhir ayat ini Allah SWT menyatakan dengan tegas bahwa tidak ada urusanpun melainkan tercatat dalam kitab yang nyata kitab *(Lawh Mahfuz)*. Maksudnya segala macam urusan itu semuanya dikontrol, dikendalikan, dan dikuasai oleh ilmu Allah Yang Maha Luas dan tercatat dalam kitab-Nya yang bernilai tinggi dan sempurna uraiannya.

Analisis regresi merupakan salah satu alat statistik yang digunakan untuk mempelajari hubungan antara variabel respon dan variabel prediktor. Ada dua metode yang dapat digunakan untuk menaksir fungsi regresi, yaitu metode regresi parametrik dan metode regresi nonparametrik. Metode regresi parametrik akan sesuai jika bentuk fungsi regresi diketahui. Tetapi jika fungsi regresi tidak diketahui bentuknya, maka regresi perlu diestimasi. Fungsi regresi dapat diestimasi, salah satunya dengan pendekatan nonparametrik. Dalam hal ini fungsi regresi hanya diasumsikan termuat dalam suatu ruang fungsi tertentu, dimana pemilihan ruang fungsi tersebut biasanya dimotivasi oleh sifat kemulusan (*smoothness*), artinya mempunyai turunan yang kontinu (Wahba, 1990). Salah satu persoalan yang sering muncul dalam pendugaan parameter regresi, yaitu tidak semua parameter dapat didekati dengan parametrik maupun nonparametrik. Model seperti ini menghasilkan model regresi semiparametrik.

Salah satu analisis regresi adalah regresi semiparametrik yang merupakan gabungan dari regresi parametrik dan nonparametrik. Kajian mengenai analisis regresi semiparametrik terinspirasi dari salah satu ayat Al-Qur'an seperti dalam surat Al-Ahzab (33) ayat 35 yang berbunyi:

إِنَّ الْمُسْلِمِينَ وَالْمُسْلِمَاتِ وَالْمُؤْمِنِينَ وَالْمُؤْمِنَاتِ وَالْقَانِتِينَ وَالْقَانِتَاتِ وَالصَّادِقِينَ وَالصَّادِقَاتِ وَالصَّابِرِينَ وَالصَّابِرَاتِ وَالْخَاشِعِينَ وَالْخَاشِعَاتِ وَالْمُتَصَدِّقِينَ وَالْمُتَصَدِّقَاتِ وَالصَّائِمِينَ وَالصَّائِمَاتِ وَالْحَافِظِينَ فُرُوجَهُمْ وَالْحَافِظَاتِ وَالذَّاكِرِينَ اللَّهَ كَثِيرًا وَالذَّاكِرَاتِ أَعَدَّ اللَّهُ لَهُمْ مَغْفِرَةً وَأَجْرًا عَظِيمًا

Artinya:

"Sesungguhnya laki-laki dan perempuan yang muslim, laki-laki dan perempuan yang mukmin, laki-laki dan perempuan yang tetap dalam ketaatannya, laki-laki dan perempuan yang benar, laki-laki dan perempuan yang sabar, laki-laki dan perempuan yang khusyuk, laki-laki dan perempuan yang bersedekah, laki-laki dan perempuan yang berpuasa, laki-laki dan perempuan yang memelihara kehormatannya, laki-laki dan perempuan yang banyak menyebut (nama) Allah, Allah telah menyediakan untuk mereka ampunan dan pahala yang besar" (QS. Al-Ahzab: 35).

Dari ayat di atas, *"Sungguh, laki-laki dan perempuan muslim, laki-laki dan perempuan mukmin"* menyatakan bahwa iman berbeda dengan Islam. Iman lebih spesifik dari pada Islam. Firman Allah SWT *"Laki-laki dan perempuan yang benar"* adalah jujur dalam ucapan. Jujur adalah sifat terpuji. Firman Allah SWT *"Laki-laki dan perempuan yang sabar"* adalah sabar dalam menghadapi musibah. Firman Allah SWT *"Laki-laki dan perempuan yang khusyuk"*, khusyuk adalah ketenangan dan ketenteraman. Orang yang khusyuk senantiasa takut kepada Allah SWT dan merasakan pengawasan-Nya. Firman Allah SWT *"laki-laki dan perempuan yang bersedekah"* sedekah adalah berbuat baik kepada yang membutuhkan dan lemah. Firman Allah SWT *"Allah telah menyediakan untuk mereka ampunan dan pahala yang besar"*. Sesungguhnya Allah SWT menyediakan bagi mereka ampunan dosa dan pahala yang besar, yaitu surga.

Inspirasi ayat di atas terkait analisis regresi semiparametrik dapat diilustrasikan sebagai berikut: variabel respon dari ayat di atas adalah ampunan dan pahala yang besar. Sedangkan variabel prediktor unsur parametriknya adalah muslim, mukmin, ketaatan, benar, sabar, khusyuk, sedekah, puasa, memelihara kehormatan, dan banyak menyebut nama Allah SWT. Variabel prediktor untuk unsur nonparametriknya adalah laki-laki dan perempuan, karena dari jenis kelamin ini tidak mempengaruhi kadar ampunan dan pahala seseorang dan dengan banyaknya menyebut nama Allah SWT merupakan faktor eksternal yang mempengaruhi kadar ampunan dan pahala seseorang.

Salah satu pendekatan regresi semiparametrik untuk memperoleh estimasi kurva regresi adalah *smoothing spline*. Masalah utama pada saat mengestimasi kurva regresi menggunakan *smoothing spline* adalah memilih dan menentukan parameter penghalus. *Smoothing spline* akan mengestimasi fungsi regresi sebagai solusi dari masalah optimasi, yaitu dengan mencari kurva penghalus dengan meminimumkan jumlah kuadrat galat terpenalti (*Penalized Residual Sum of Square*) yang terdiri atas kuadrat tengah sisaan *Mean Square Error (MSE*) dan parameter penghalus yang merupakan penalti kekasaran (*roughness penalty*) yang memberi ukuran kemulusan atau kekasaran kurva dalam memetakan data. Parameter penghalus mengendalikan perimbangan (*trade off*) antara ketepatan model (*goodness of fit*) dan mulusnya estimator.

Metode pemilihan parameter penghalus dibagi menjadi dua metode, yaitu metode klasik dan metode estimasi resiko. Metode klasik terdiri atas *Cross-Validation* (*CV*), *Generalized Cross-Validation* (GCV), kriteria Mallows' *Cp* dan *Improved Akaike Information Criterion* (*AIC*). Sedangkan metode estimasi resiko terdiri atas *Risk Estimation Using Classical Pilots* (*RECP*) dan *Exact Double Smoothing* (*EDS*).

Regresi semiparametrik dengan pendekatan *smoothing spline* pada hakekatnya merupakan generalisasi dari fungsi polinom biasa, dimana cara-cara mengoptimasinya masih mengadopsi konsep dalam regresi parametrik.

Pendekatan *smoothing spline* dapat dilakukan dengan bantuan pemilihan parameter penghalus optimal dengan menggunakan metode *Generalized Cross Validation* (GCV).

Regresi semiparametrik dapat diaplikasikan ke dalam kehidupan sehari-hari. Pada penelitian ini yang mengaplikasikan regresi semiparametrik dalam kasus kemiskinan di Kota Malang tahun 2005-2012. Di samping itu kemiskinan merupakan hal yang sangat penting diperhatikan dan perlu dikurangi untuk meningkatkan kesejahteraan hidup suatu rumah tangga. Banyak faktor yang mempengaruhi jumlah penduduk miskin, baik faktor internal maupun faktor eksternal. Faktor internal biasanya merupakan variabel prediktor untuk komponen parametrik dan faktor eksternal merupakan variabel nonparametrik karena hanya memiliki pengaruh yang kecil terhadap varibel respon. Sehingga regresi nonparametrik diasumsikan mulus karena berperan sebagai *error* untuk memuluskan fungsi. Dalam regresi semiparametrik yang merupakan gabungan dari parametrik dan nonparametrik dimana variabel prediktor unsur nonparametriknya berperan sebagai pemulus untuk memberikan ketepatan model. Menggabungkan faktor internal dan faktor eksternal yang mempengaruhi suatu variabel respon sehingga terbentuk suatu model yang tidak hanya dilihat dari faktor internal saja dan akan terlihat seberapa pengaruh dari masing-masing variabel prediktor terhadap variabel respon.

Beberapa studi tentang kemiskinan sudah banyak dilakukan oleh beberapa peneliti dari beberapa negara. Sebagian besar kemiskinan di beberapa negara disebabkan oleh pengangguran dan keterbelakangan yang menyebabkan adanya ketimpangan sosial serta rendahnya tingkat pendidikan. Masyarakat miskin pada umumnya lemah dalam kemampuan berusaha dan terbatasnya akses dalam kegiatan ekonomi sehingga membuat mereka semakin miskin. Akan tetapi seburuk apapun tingkat kesejahteraan ekonomi pasti ada solusinya. Allah SWT berfirman dalam surat Al Isra' (17) ayat 31:

وَلَا تَقْتُلُوا أَوْلَادَكُمْ خَشْيَةَ إِمْلَاقٍ ۖ نَحْنُ نَرْزُقُهُمْ وَإِيَّاكُمْ ۚ إِنَّ قَتْلَهُمْ كَانَ خِطْئًا كَبِيرًا

Artinya:
"*Dan janganlah kamu membunuh anak-anakmu karena takut kemiskinan. Kamilah yang akan memberi rezki kepada mereka dan juga kepadamu. Sesungguhnya membunuh mereka adalah suatu dosa yang besar*" *(QS. Al Isra': 31).*

Ayat di atas menerangkan bahwa Allah melarang manusia untuk membunuh anak keturunannya, dikarenakan takut akan kemiskinan. Allah SWT menjamin rezki setiap hamba-Nya, setiap manusia dan semua makhluk Allah SWT yang lahir ke dunia telah dipersiapkan rezkinya. Namun demikian, rezki didapat melalui ikhtiar (usaha), Allah memerintahkan kepada manusia untuk bekerja jika mereka ingin memenuhi kebutuhan hidupnya, seperti kebutuhan akan makanan dan minuman.

Sebagai referensi pada penelitian ini, beberapa jurnal yang mengkaji tentang parameter penghalus *spline* yang telah dilakukan oleh Thomas C.M. Lee (2003) yang menyimpulkan bahwa metode *AIC* tidak pernah memberikan hasil yang buruk dibanding metode *CV*, GCV dan *Cp* namun untuk fungsi yang sederhana dengan tingkat *noise* yang tinggi, metode estimasi resiko terlihat lebih unggul. Sedangkan penelitian yang dilakukan oleh Dursun Aydin dan M.S. Tuzemen (2012) tentang masalah pemilihan parameter penghalus dalam regresi nonparametrik berdasarkan *smoothing spline* menyimpulkan bahwa khusus untuk sampel besar paling baik menggunakan metode *AIC*, sedangkan khusus untuk sampel kecil paling baik menggunakan metode GCV. Berdasarkan latar belakang tersebut penulis ingin membahas metode pemilihan parameter penghalus GCV pada regresi semiparametrik sehingga penulis mengambil judul *"Penentuan Parameter Penghalus Smoothing Spline dalam Regresi Semiparametrik dengan GCV"*, dimana dalam penelitian ini akan dicari parameter penghalus optimal yang menggunakan metode GCV dari model regresi semiparametrik, selanjutnya akan diaplikasikan pada data penelitian.

1.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini antara lain:

1. Bagaimana menentukan parameter penghalus optimal pada *smoothing spline* dengan metode GCV?
2. Bagaimana model regresi semiparametrik *smoothing spline* pada tingkat kemiskinan di Kota Malang tahun 2005-2012?

1.3 Tujuan Penelitian

Tujuan penelitian ini sebagai berikut:

1. Menentukan parameter penghalus optimal pada *smoothing spline* dengan metode GCV
2. Mengetahui model regresi semiparametrik *smoothing spline* pada tingkat kemiskinan di Kota Malang tahun 2005-2012

1.4 Batasan Masalah

Batasan masalah dalam penelitian ini adalah

1. Estimasi parameter pada model regresi semiparametrik menggunakan pendekatan regresi *smoothing spline cubic* dan optimasi *Penalized Least Square (PLS)*
2. Nilai parameter α pada regresi *smoothing spline cubic* dalam metode Newton Raphson ditentukan antara 0.01-0.05 untuk memperkecil tingkat signifikan atau kesalahan yang ditoleransi dalam membuat keputusan
3. Nilai λ diberikan $0 < \lambda < 1$ yang ditentukan secara acak menggunakan program MATLAB
4. Data semiparametrik yang digunakan adalah data jumlah kemiskinan di Kota Malang pada tahun 2005-2012 sebagai variabel respon, data tingkat pengangguran terbuka sebagai variabel prediktor unsur parametrik dan indeks pendidikan sebagai variabel prediktor unsur noparametriknya

1.5 Manfaat Penelitian

Adapun manfaat yang dapat diperoleh dari penelitian ini, yaitu:

1. Bagi penulis

   Penelitian ini digunakan sebagai tambahan informasi dan wawasan tentang parameter penghalus dari *smoothing spline* pada regresi semiparametrik dan mengaplikasikannya dalam kehidupan sehari-hari.

2. Bagi mahasiswa matematika

Sebagai motivasi agar bisa mengembangkan dan menerapkan ilmu matematika ke dalam bidang keilmuan lain.

3. Bagi pembaca

Sebagai tambahan wawasan dan informasi tentang aplikasi pengembangan ilmu matematika dalam bidang keilmuan lain.

1.6 Metode Penelitian

1.6.1 Pendekatan Penelitian

Dalam penelitian ini menggunakan pendekatan *library research*, dimana dalam pendekatan *library research* ini dikaji secara literatur yang diambil dari buku pustaka dan artikel ilmiah yang diunduh dari sumber internet.

1.6.2 Langkah-langkah Penelitian

1.6.2.1 Menentukan Parameter Penghalus Optimal pada Model Regresi *Smoothing Spline* dengan Metode GCV

Untuk menyelesaikan penelitian dalam skripsi ini, penulis membuat langkah-langkah dalam menentukan parameter penghalus optimal sebagai berikut:

1. Mengkaji *Penalized Least Square (PLS)*
2. Menurunkan *PLS* terhadap β, $\frac{\partial PLS}{\partial \beta} = 0$
3. Menurunkan *PLS* terhadap f, $\frac{\partial PLS}{\partial f} = 0$
4. Menentukan $\hat{\beta}_\lambda$ dan $\hat{f}_\lambda$ dengan mensubtitusikan hasil dari point 1 dan 2

5. Menentukan *hat matrix* $A(\lambda)$ dengan mensubtitusikan $\hat{\beta}_\lambda$ dan $\hat{f}_\lambda$ ke dalam $\hat{Y} = X\hat{\beta} + \hat{f}$
6. Menentukan parameter penghalus optimal dengan mensubtitusikan *hat matrix* ke dalam fungsi GCV

1.6.2.2 Analisis Data

Untuk melengkapi penelitian dalam skripsi, penulis menganalisis data kemiskinan di Kota Malang tahun 2005-2012 dengan langkah-langkah sebagai berikut:

1. Mendeskripsi data kemiskinan di Kota Malang tahun 2005-2012
2. Mendeteksi variabel prediktor komponen parametrik (X) dan variabel prediktor komponen nonparametrik (t) menggunakan *scatter plot* menggunakan program minitab 14
3. Memilih model tentatif yang terbaik untuk *smoothing spline cubic* menggunakan program MATLAB
4. Menguji signifikansi model regresi semiparametrik

1.7 Sistematika Penulisan

Untuk lebih mudah memahami penulisan ini secara keseluruhan isinya, maka penulis memberikan gambaran umum tentang sistematika penulisan sebagai berikut:

Bab I Pendahuluan

Pada bab pertama ini dibahas tentang latar belakang penelitian, rumusan masalah, tujuan penelitian, batasan masalah, manfaat penelitian, metode penelitian, dan sistematika penulisan.

Bab II Tinjauan Pustaka

Pada bab kedua ini akan dibahas beberapa teori yang ada kaitannya dengan hal-hal yang penulis bahas.

Bab III Pembahasan

Pada bab ketiga ini dibahas tentang proses agar memperoleh nilai fungsi parameter penghalus dari model regresi *smoothing spline* dan nilai parameter penghalus dari model regresi tersebut serta aplikasinya dalam data..

Bab IV Penutup

Pada bab keempat ini berisi tentang kesimpulan dari pembahasan berdasarkan rumusan masalah dan saran yang berkaitan dengan penulisan. Saran ini diharapkan dapat memberikan masukan yang positif untuk dikembangkan.

BAB II

TINJAUAN PUSTAKA

2.1 Regresi Parametrik

Regresi parametrik merupakan metode statistika yang digunakan untuk mengestimasi bentuk hubungan antara variabel prediktor dengan variabel respon dimana bentuk kurva regresinya diketahui. Hubungan antara variabel respon dan prediktor dalam model dapat terjadi dengan fungsi linier maupun nonlinier dalam parameter.

Secara umum model regresi parametrik dengan satu variabel prediktor adalah

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \,,\; i = 1, 2, \ldots, n \qquad (2.1)$$

$$Y = X\beta + \varepsilon \qquad (2.2)$$

dimana

y_i : variabel respon dari data ke-i

β_0, β_1 : parameter yang tidak diketahui yang akan diestimasi

x_i : variabel prediktor dari data ke-i

ε_i : *error* ke-*i* yang diasumsikan menyebar $N \sim (0, \sigma^2)$ (Ruppert, 2003).

2.2 Regresi Nonparametrik

Regresi nonparametrik merupakan salah satu metode statistik yang digunakan untuk mengestimasi bentuk hubungan antara variabel respon dengan variabel prediktor tidak diketahui bentuk kurva regresinya. Dalam regresi nonparametrik kurva regresi hanya diasumsikan mulus (*smooth*) dalam arti termuat dalam suatu ruang fungsi tertentu sehingga mempunyai fleksibilitas yang tinggi (Eubank, 1999).

Secara umum model regresi nonparametrik dapat dituliskan sebagai berikut:

$$y_i = f\left(t_i\right) + \varepsilon_i\,,\; i = 1, 2, \ldots, n \tag{2.3}$$

$$Y = f + \varepsilon \tag{2.4}$$

dimana

y_i : variabel respon dari data ke- i

$f(t_i)$: fungsi regresi yang tidak diketahui bentuknya (fungsi mulus)

t_i : variabel prediktor dari data ke- i

ε_i : *error* ke- i yang diasumsikan menyebar $N\sim(0,\sigma^2)$ (Wahba, 1990).

2.3 Regresi Semiparametrik

Model regresi semiparametrik merupakan bentuk pola hubungan antara variabel respon dan variabel prediktor yang terdiri dari regresi parametrik dan regresi nonparametrik. Memiliki kedua komponen parametrik dan nonparametrik berarti model semiparametrik. Kelas ini model semiparametrik sederhana adalah penting dalam dirinya sendiri tetapi juga berfungsi sebagai pengantar regresi semiparametrik lebih kompleks dan efek dari beberapa prediktor dimodelkan secara nonparametrik. Sehingga model regresi semiparametrik adalah

$$y_i = \beta_0 + \beta_1 x_i + f\left(t_i\right) + \varepsilon_i\,,\; i = 1,2,\ldots,n \tag{2.5}$$

$$Y = X\beta + f + \varepsilon \tag{2.6}$$

dimana

y_i : variabel respon dari data ke-i

β_0, β_1 : parameter yang tidak diketahui yang akan diestimasi

x_i : variabel prediktor dari data ke-i untuk kompenen parametrik

$f(t_i)$: fungsi regresi yang tidak diketahui bentuknya (fungsi mulus) untuk komponen nonparametrik

t_i : variabel prediktor dari data ke-i

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

Model regresi semiparametrik terdiri dari unsur parametric $\beta_0 + \beta_1 x_i$ dan unsur nonparametrik $f(t_i)$ (Ruppert, 2003).

2.4 Estimasi Parameter

Dalam statistik, estimasi adalah metode yang digunakan untuk mengetahui nilai-nilai populasi penaksiran dengan nilai-nilai sampel. Nilai-nilai populasi juga sering disebut populasi parameter. Mengingat nilai-nilai sampel disebut sampel statistik. Dalam metode estimasi, populasi parameter yang diestimasi adalah rata-rata yang ditulis dengan notasi p dan standar deviasi yang ditulis dengan notasi q. Dengan nilai-nilai sampel, penulis mencoba untuk mengetahui populasi karakteristik.

Estimasi adalah suatu proses yang menggunakan sampel (statistik) untuk mengetahui hubungan parameter dengan populasi yang belum diketahui. Estimasi adalah pernyataan mengenai populasi parameter yang diketahui dengan sampel.

Dalam kasus ini, variabel random diambil dari hubungan populasi. Dengan estimasi ini, kondisi populasi parameter yang diketahui (Hasan, 2001).

2.4.1 Metode *Least Square*

Untuk mendapatkan estimator dari kuadrat *error* model regresi kemudian menurunkannya. Hal ini bertujuan untuk mendapatkan nilai β dengan meminimalkan jumlah kuadrat. Dengan menggunakan metode *Least Square*, parameter β dalam model regresi (2.2) dapat diestimasi dengan meminimumkan $\varepsilon^T \varepsilon$ terhadap β. Untuk $\varepsilon^T \varepsilon = (Y - X\beta)^T (Y - X\beta)$, dengan menurunkan $\varepsilon^T \varepsilon$ terhadap β dan membuat $\frac{\partial \varepsilon^T \varepsilon}{\partial \beta} = 0$, maka diperoleh estimator

$$\hat{\beta} = \left(X^T X\right)^{-1} X^T Y \tag{2.7}$$

dengan syarat $X^T X$ mempunyai invers (Ruppert, 2003).

Berikut langkah untuk menentukan estimasi parameter dari β pada unsur parametrik dengan metode *Least Square*:

$$Y = X\beta + \varepsilon$$

$$\varepsilon = Y - X\beta \tag{2.8}$$

$$\varepsilon^T \varepsilon = (Y - X\beta)^T (Y - X\beta) \tag{2.9}$$

$$= \left(Y^T - \beta^T X^T\right)(Y - X\beta)$$

$$= Y^T Y - \beta^T X^T Y - Y^T X\beta + \beta^T X^T X\beta \tag{2.10}$$

Kemudian, meminimalkan persamaan (2.10) dengan menurunkan terhadap β sama dengan nol.

$$\frac{\partial \varepsilon^T \varepsilon}{\partial \beta} = 0$$

$$\frac{\partial\left(Y^TY-\beta^TX^TY-Y^TX\beta+\beta^TX^TX\beta\right)}{\partial\beta}=0$$

$$-X^TY-\left(Y^TX\right)^T+X^TX\beta+\left(\beta^TX^TX\right)^T=0$$

$$-X^TY-X^TY+X^TX\beta+X^TX\beta=0$$

$$-2X^TY+2X^TX\beta=0$$

$$2X^TX\beta=2X^TY$$

$$X^TX\beta=X^TY$$

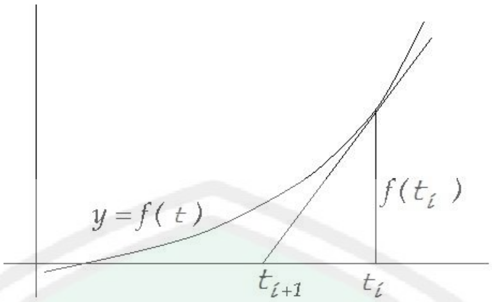
$$\left(X^TX\right)^{-1}\left(X^TX\right)\beta=\left(X^TX\right)^{-1}X^TY$$

$$I\beta=\left(X^TX\right)^{-1}X^TY$$

$$\hat{\beta}=\left(X^TX\right)^{-1}X^TY \tag{2.11}$$

2.4.1 Metode Newton Raphson

Metode Newton Raphson adalah metode yang digunakan untuk menemukan akar dari persamaan nonlinier karena konvergensinya yang sangat cepat dibandingkan dengan metode lainnya. Misalkan t_i adalah aproksimasi untuk akar α dari $f(t)=0$ yang ditunjukkan pada gambar (2.1). Untuk aproksimasi selanjutnya yaitu t_{i+1} diambil dari titik pada garis singgung dari grafik $y_i = f(t_i)$ pada $t=t_i$ yang bertemu di sumbu t. Misalkan persamaan dari garis singgung itu adalah $y=zt+d$ dan $f'(t_n)$ adalah lengkung atau kemiringan dari garis singgung z pada $t=t_i$. Jika garis singgung melewati titik $(t_i, f(t_i))$, maka grafik persamaan $y=zt+d$ menjadi gambar (2.1):



Gambar 2.1Grafik Newton Raphson

$$f(t_i) = f'(t_i)t_i + d \tag{2.12}$$

dimana

$$d = f(t_i) - f'(t_i)t_i \tag{2.13}$$

dengan memasukkan nilai d, persamaan garis singgungnya menjadi:

$$y = f'(t_i)t_i + f(t_i) - f'(t_i)t_i \tag{2.14}$$

Untuk aproksimasi selanjutnya, yaitu t_{i+1} adalah titik potong atau titik pertemuan dari sumbu t dan garis singgung dari $t = t_i$. Kemudian, dengan memasukkan nilai $y = 0$ dalam persamaan (2.14), diperoleh:

$$0 = f'(t_i)t_{i+1} + f(t_i) - f'(t_i)t_i$$

sehingga

$$t_{i+1} = \frac{f'(t_i)t_i - f(t_i)}{f'(t_i)}, \quad f'(t_i) \neq 0 \tag{2.15}$$

dimana persamaan ini disebut Newton Raphson (Naseem, 2010).

Berikut contoh estimasi regresi nonparametrik menggunakan metode Newton Raphson:

Dari model regresi nonparametrik (2.3), $y_i = f(t_i) + \varepsilon_i$, $i = 1, 2, ..., n$

dimana fungsi $f(t_i)$ belum diketahui, maka untuk mengestimasinya menggunakan metode Newton Raphson persamaan (2.14):

$$y = f'(t_i)t_i + f(t_i) - f'(t_i)t_i$$

Untuk aproksimasi selanjutnya, yaitu t_{i+1} adalah titik potong atau titik pertemuan dari sumbu t dan garis singgung dari $t = t_i$. Kemudian, dengan memasukkan nilai $y = 0$ dalam persamaan (2.14), diperoleh:

$$0 = f'(t_i)t_{i+1} + f(t_i) - f'(t_i)t_i$$

$$f'(t_i)t_{i+1} = f'(t_i)t_i - f(t_i)$$

$$t_{i+1} = \frac{f'(t_i)t_i - f(t_i)}{f'(t_i)}$$

$$t_{i+1} = t_i - \frac{f(t_i)}{f'(t_i)},\ f'(t_i) \neq 0 \tag{2.16}$$

sehingga estimasi fungsi $f(t_i)$ dari Newton Raphson adalah

$$f(t_i) = f'(t_i)t_i - f'(t_i)t_{i+1}$$

$$\hat{f}(t_i) = (t_i - t_{i+1})\, f'(t_i) \tag{2.17}$$

dimana $f(t_i)$ adalah persamaan *spline cubic* (2.23):

$$f(t_i) = \alpha_0 + \alpha_1 t_i + \alpha_2 t_i^2 + \alpha_3 t_i^3 + \sum_{j=1}^{s} u_j (t_i - k_j)_+^3$$

yang akan diestimasi menggunakan metode Newton Raphson dengan bantuan *software* MATLAB. Nilai α diberikan secara acak $0,01$ sampai $0,05$ untuk memperkecil tingkat signifikan atau kesalahan yang ditolerir dalam membuat keputusan dan titik knot diberikan secara acak dengan $k_1 < k_2 < \ldots < k_s$ yang akan menghasilkan iterasi paling cepat. Hasil dari perhitungan dalam program, $f\left(t_i\right)$ yang mendapat iterasi paling cepat adalah sebagai berikut:

$$f(t)=0.01+0.02t+0.03t^2+0.04t^3+0.01(t-3)^3+0.02(t-4)^3+0.03(t-5)^3+ 0.04\left(t-6\right)^3 \quad (2.18)$$

dengan hasil turunan pertama dari fungsi $f(t_i)$ adalah sebagai berikut:

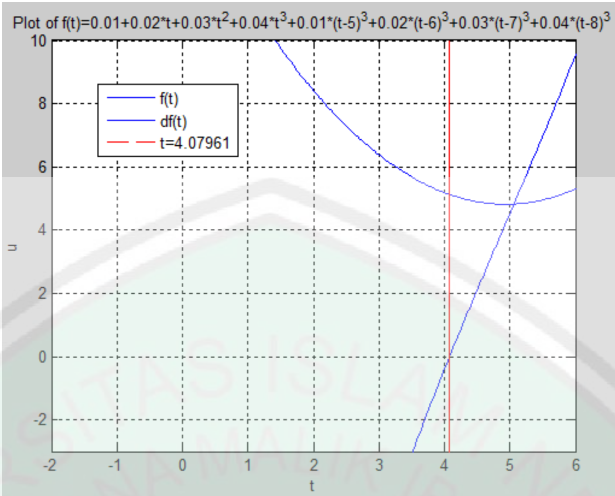
$$f'\left(t\right)=0.2e-1+0.6e-1*t+0.12*t^2+0.3e-1*(t-3)^2+0.6e-1*(t-4)^2+ 0.9e-1*\left(t-5\right)^2+0.12*\left(t-6\right)^2 \quad (2.19)$$

dimana

$$t_{i+1}=t_i-\frac{f(t_i)}{f'(t_i)}$$

sehingga dihasilkan nilai $t_{i+1}=4.079614157068315$.

Dalam perhitungan program di atas, nilai α diberikan random dengan $\alpha_0=0.01$, $\alpha_1=0.02$, $\alpha_2=0.03$, $\alpha_3=0.04$, $\alpha_4=0.01$, $\alpha_5=0.02$, $\alpha_6=0.03$ dan $\alpha_7=0.04$. Sedangkan titik knot diberikan $k_1=5$, $k_2=6$, $k_3=7$ dan $k_4=8$. Dengan diberikan nilai awal $t_0=3$. Nilai t_{i+1} berhenti pada iterasi keempat dengan nilai 4.079614157068315. Dan nilai turunan pertamanya $f'\left(t_{i+1}\right)=5.120563091369107$.



Gambar 2.2 Grafik Newton Raphson Fungsi Nonparametrik

Dari gambar 2.2 di atas fungsi $f(t)$ berupa kurva yang lebih cenderung lurus dan turunan pertama fungsi $f(t)$, yaitu $f'(t)$ berupa kurva yang menghadap ke atas. Untuk iterasi tercepat dari $f(t)$ jatuh pada iterasi keempat yang mana memotong sumbu $y=0$.

Sehingga untuk mendapatkan estimasi fungsi semiparametrik, yaitu dengan mensubtitusikan estimator $\hat{\beta}$ dan $\hat{f}$ ke dalam persamaan berikut:

$$\begin{aligned}\hat{Y} &= X\hat{\beta}+\hat{f} \\ &= X\left[\left(X^TX\right)^{-1}X^TY\right]+\hat{f}(t_i) \\ &= X\left(X^TX\right)^{-1}X^TY+\left(t_i-t_{i+1}\right)f'(t_i) \qquad (2.20)\end{aligned}$$

2.5 Asumsi-asumsi Analisis Regresi Semiparametrik

Adapun asumsi yang digunakan pada regresi semiparametrik, yaitu:

a. $\varepsilon \backsim N(0,\sigma^2)$ dengan model acak yang saling independen dengan mean 0 dan varian σ^2

b. Antara variabel prediktor adalah saling bebas (non-multikolinieritas)

c. Sebagian bentuk kurva diketahui dan sebagian bentuk kurva tidak diketahui (Ruppert, 2003).

2.6 Uji Asumsi pada *Error*

2.6.1 Uji Asumsi Kenormalan

Dalam analisis regresi diperlukan pengujian terhadap normalitas pada *error*. Untuk uji normalitas bisa juga digunakan uji *Jarque-Bera* (*JB*) pengujian normalitas dengan uji *Jarque-Bera* (*JB*) mengunakan formula sebagai berikut:

$$JB = n\left[\frac{S^2}{6} + \frac{(K-3)^2}{24}\right] \tag{2.21}$$

dimana *S* menunjukkan *Skewness* dan *K* menunjukkan *Kurtosis.* Kesalahan pengganggu kemungkinan berasal dari distribusi normal jika nilai JB lebih kecil dari nilai $\chi^2_{\alpha,df}$ tertentu (Algifari, 2000).

Skewness bisa didapatkan dari hasil bagi momen ketiga rata-rata dengan pangkat tiga dari standar deviasi, sedangkan *Kurtosis* bisa diperoleh dari hasil bagi rata-rata momen keempat dengan kuadrat dari momen kedua, sehingga bisa dirumuskan:

$$S = \frac{E\left(X - E(X)\right)^3}{\sigma^3}$$

dan

$$S = \frac{E\left(X - E(X)\right)^4}{\left[E\left(X - E(X)\right)^2\right]^2} \qquad \text{(Anonim, 2013)}$$

2.6.2 Uji Asumsi Non-multikolinearitas

Istilah multikolinearitas diciptakan oleh Ragner Frish. Istilah itu berarti adanya hubungan linier yang sempurna atau eksak di antara variabel-variabel bebas dalam model regresi. Apabila terjadi kolinearitas sempurna maka koefisien regresi dari variabel X tidak dapat ditentukan dan standar *error*-nya tinggi, yang berarti koefisien regresi tidak dapat diperkirakan dengan tingkat ketelitian yang tinggi. Jadi semakin kecil korelasi di antara variabel bebasnya maka semakin baik model regresi yang akan diperoleh. Dengan demikian, masalah multikolinearitas adalah masalah derajat (Firdaus, 2004).

Kolinearitas seringkali dapat diestimasi jika nilai R^2 cukup tinggi dan koefisien regresi sederhana juga tinggi. Akan tetapi tidak satupun atau sedikit sekali koefisien regresi parsial yang signifikan secara individu jika dilakukan *uji-t*, maksudnya hipotesis nol bahwa koefisien regresi parsial sama dengan nol hampir semuanya diterima. Jadi secara individual tidak mempunyai pengaruh terhadap variabel bebas Y. Apabila nilai R^2 tinggi, ini berarti bahwa *uji-F* melalui analisis varian, pada umumnya akan menolak hipotesis nol yang mengatakan bahwa secara simultan atau bersama-sama, koefisien regresi parsialnya nol (Supranto, 2005).

2.6.3 Uji Asumsi Non-autokorelasi

Salah satu asumsi pada model regresi semiparametrik adalah non-autokorelasi. Tetapi asumsi tersebut tidak selalu dipenuhi, jika gangguan penyimpangan berupa autokorelasi secara nyata ada pada suatu fungsi regresi maka asumsi tersebut tidak berlaku. Berarti pada fungsi regresi tersebut ada autokorelasi. Autokorelasi merupakan gangguan pada fungsi regresi yang berupa korelasi di antara faktor pengganggu.

Sebagai akibatnya autokorelasi pada model persamaan regresi maka akan terjadi hal-hal berikut:

1. Estimasi-estimasi koefisien regresi yang diperoleh tetap merupakan estimasi yang tak bias
2. Varian variabel pengganggu menjadi tidak efisien jika dibandingkan dengan tidak adanya autokorelasi. Varian variabel pengganggu mungkin sekali akan dinilai terlalu rendah, sehingga akibatnya uji statistik yang digunakan terhadap koefisien estimasi regresi berkurang kemaknaannya dan mungkin akan menjadi tidak berarti sama sekali

Ada beberapa prosedur untuk mengetahui adanya masalah autokorelasi pada suatu model regresi. Tetapi uji ada tidaknya autokorelasi yang paling banyak digunakan adalah Uji Durbin Watson (uji *DW*). Uji ini dapat digunakan untuk sembarang sampel baik besar atau kecil, tetapi uji *DW* hanya berhasil baik apabila autokorelasinya berbentuk autokorelasi linier orde pertama.

Hipotesis: $H_0 : \rho = 0$, tidak ada autokorelasi antar *error*

$H_1 : \rho \neq 0$, ada autokorelasi antar *error*

Rumus besarnya nilai statistik *DW*:

$$DW = \frac{\sum_{2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} \tag{2.22}$$

dimana

DW : statistik uji Durbin Watson

e_{t-1} : kesalahan pengganggu pada periode $t-1$

e_t : kesalahan pengganggu pada periode t

Ketentuan untuk melihat ada tidaknya autokorelasi diberikan sebagai berikut:

Tabel 2.1 Ketentuan Durbin Watson

DW	Kesimpulan
<1.10	Ada autokorelasi
1.10-1.54	Tanpa kesimpulan
1.55-2.45	Tidak ada autokorelasi
2.46-2.90	Tanpa kesimpulan
>2.91	Ada autokorelasi

(Sumber: Firdaus, 2004)

2.6.4 Uji Asumsi Kehomogenan Ragam *Error*

Salah satu asumsi pada *error* adalah var $(\varepsilon_i) = \sigma^2$, yaitu variasi dari faktor pengganggu selalu sama pada data pengamatan yang satu ke data pengamatan yang lain. Jika hal itu dipenuhi berarti variasi faktor pengganggu pada kelompok data tersebut bersifat homokedastik. Jika asumsi tersebut tidak dipenuhi maka terjadi penyimpangan terhadap faktor pengganggu yang disebut homoskedastisitas (Firdaus, 2004).

2.7 *P-Value*

Dalam ilmu statistika, para peneliti harus menggunakan kriteria uji untuk memutuskan apakah menolak H_0 atau menerima H_0. Dalam perkembangannya, banyak peneliti yang sering menggunakan *p-value* untuk kriteria ujinya. *P-value* lebih disukai dibandingkan kriteria uji lain seperti tabel distribusi dan selang kepercayaan. Hal ini disebabkan karena *p-value* memberikan dua informasi sekaligus, yaitu di samping petunjuk apakah H_0 pantas ditolak, *p-value* juga memberikan informasi mengenai peluang terjadinya kejadian yang disebutkan di dalam H_0 (dengan asumsi H_0 dianggap benar). Definisi *p-value* adalah tingkat keberartian terkecil sehingga nilai suatu uji statistik yang sedang diamati masih berarti. *P-value* dapat juga diartikan sebagai besarnya peluang melakukan kesalahan apabila kita memutuskan untuk menolak H_0 (Kurniawan, 2008). Pada umumnya, *p-value* dibandingkan dengan suatu taraf nyata α tertentu, biasanya 0.05 atau 5%. Taraf nyata α diartikan sebagai peluang dalam melakukan kesalahan untuk menyimpulkan bahwa H_0 salah, padahal sebenarnya *statement* H_0 yang benar. Kesalahan semacam ini biasa dikenal dengan galat/kesalahan jenis 1.

2.8 Kelayakan Model

Koefisien determinasi yang disimbolkan dengan R^2 adalah alat untuk mengukur proporsi keragaman atau variansi total di sekitar nilai tengah y yang dapat dijelaskan oleh model regresi. Secara umum semakin besar nilai R^2, maka semakin baik pula model yang didapatkan karena mampu menjelaskan lebih banyak data, dengan rumus koefisien determinasi:

$$R^2 = \frac{\sum\left(\hat{y}_i - \bar{y}\right)^2}{\sum\left(y_i - \bar{y}\right)^2} \qquad (2.23)$$

dimana

R^2 : koefisien determinasi

$\hat{y}$: nilai estimasi peubah respon ke–*i*

$\bar{y}$: rata-rata peubah respon

y_i : nilai peubah respon ke–*i* (Draper dan Smith, 1998).

2.9 *Spline*

Spline merupakan potongan polinomial *(piecewise polynomial)* tersegmen yang memiliki sifat fleksibilitas. Titik perpaduan bersama dari potongan-potongan tersebut atau titik yang menunjukkan terjadinya perubahan-perubahan perilaku kurva pada interval-interval yang berbeda disebut knot (Fan dan Yao, 2005).

Secara umum fungsi *spline* dengan orde m dengan titik-titik knot $k_1, k_2, \ldots, k_s$ untuk substansi $k \le n$ (mean $k < n$) dan jumlah prediktor dapat ditulis ke dalam bentuk $1, t, \ldots, t^m, (t-k_1)_+^m, \ldots, (t-k_s)_+^m$. Regresi *spline* dengan orde $m = 0,1,2,$ dan 3 disebut *constant*, *linier*, *quadratic*, dan *cubic*. Untuk *spline* berorde m memiliki turunan $m-1$ yang kontinu. Adapun model dari regresi *spline* adalah:

$$f(t_i) = \alpha_0 + \alpha_1 t_i + \alpha_2 t_i^2 + \ldots + \alpha_m t_i^m + \alpha_{m1}(t_i - k_1)_+^m + \ldots + \alpha_{ms}(t_i - k_s)_+^m \quad (2.24)$$

dengan $i = 1, 2, 3, \ldots, n$.

Persamaan (2.24) dapat ditulis menjadi:

$$f(t_i) = \alpha_0 + \alpha_1 t_i + \alpha_2 t_i^2 + \ldots + \alpha_m t_i^m + \sum_{j=1}^{s} u_j (t_i - k_j)_+^m \quad (2.25)$$

$r = 1, 2, 3, \ldots, m$ dan $j = 1, 2, 3, \ldots, s$, dengan

$$(t_i - k_j)_+^m = f(t_i) = \begin{cases} (t_i - k_j)^m ; t_i \ge k_j \\ 0; t_i < k_j \end{cases}$$

dimana

$f(t_i)$: fungsi regresi *spline* dari data ke-i

t_i : variabel prediktor dari data ke-i

α_0 : konstanta

α_i dan u_j : koefisien pada variabel prediktor t

$k_1, k_2, \ldots, k_s$: titik knot (Eubank, 1999).

2.10 *Spline Cubic*

Misalkan dalam persamaan (2.24) diambil orde $m = 3$, maka diperoleh sebuah *spline* yang disebut *spline cubic*. Dengan diberikan bilangan riil $(t_1, t_2, ..., t_n)$ pada suatu interval $[a,b]$ yang memenuhi $a < t_1 < t_2 < ... < t_n < b$, maka fungsi f yang terdefinisi dalam interval $[a,b]$ dikatakan *spline cubic*. Berikut adalah pendekatan model dari (2.24) berorde ke-$m = 3$ dengan satu variabel prediktor:

$$f(t_i) = \alpha_0 + \alpha_1 t_i + \alpha_2 t_i^2 + \alpha_3 t_i^3 + \alpha_{31}(t_i - k_1)_+^3 + \alpha_{32}(t_i - k_2)_+^3 + ... + \alpha_{3s}(t_i - k_s)_+^3 \tag{2.26}$$

$$f(t_i) = \alpha_0 + \alpha_1 t_i + \alpha_2 t_i^2 + \alpha_3 t_i^3 + \sum_{j=1}^{s} u_j (t_i - k_j)_+^3 \tag{2.27}$$

$f(t_i)$ adalah *spline cubic* dengan titik knot $k_1 < k_2 < ... < k_s$ dan memiliki turunan kedua yang kontinu (Eubank, 1999).

2.11 Mencari Titik Knot

Knot diartikan sebagai titik fokus dalam fungsi *spline*, sehingga kurva yang dibentuk tersegmen pada titik tersebut. Untuk mencari lokasi titik knot bisa menggunakan rumus $k_s = \left(\dfrac{k+1}{s+2}\right)$, yaitu dari banyaknya sampel yang telah diurutkan dimana $k = 1, 2, \ldots, s$. Tetapi lebih sederhana lagi memilih s dengan menggunakan rumus $s = \min\left(\dfrac{1}{4} \times number\ of\ unique\ x_i\right)$. Sebagai alternatifnya bisa juga memilih s berdasarkan diagram pencar (Ruppert, 2003).

Bentuk estimator regresi *spline* sangat dipengaruhi oleh parameter penghalus, lokasi titik knot dan banyaknya titik-titik knot. Pada hakekatnya, pemilihan parameter penghalus sama dengan memilih titik knot. Pemilihan titik knot yang optimal terletak pada nilai dan yang minimum (Budiantara, 2000). Terdapat $n-2$ kemungkinan banyaknya titik knot pada pemodelan regresi *spline* (Budiantara, 2005).

2.12 Model *Smoothing Spline* dalam Regresi Semiparametrik

Dengan model regresi *spline cubic* (2.26), yaitu:

$$f(t_i) = \alpha_0 + \alpha_1 t_i + \alpha_2 t_i^2 + \alpha_3 t_i^3 + \alpha_{31}(t_i - k_1)_+^3 + \alpha_{32}(t_i - k_2)_+^3 + \ldots + \alpha_{3s}(t_i - k_s)_+^3$$

Sehingga menjadi persamaan (2.27):

$$f(t_i) = \alpha_0 + \alpha_1 t_i + \alpha_2 t_i^2 + \alpha_3 t_i^3 + \sum_{j=1}^{s} u_j (t_i - k_j)_+^3$$

$r = 1, 2, 3, \ldots, m$ dan $j = 1, 2, 3, \ldots, s$, dengan

$$\left(t_i - k_j\right)_+^3 = f\left(t_i\right) = \begin{cases} \left(t_i - k_j\right)^3; t_i \ge k_j \\ 0; t_i < k_j \end{cases}$$

dimana

$f\left(t_i\right)$	: fungsi regresi *spline* dari data ke- i
t_i	: variabel prediktor dari data ke- i
α_0	: konstanta
α_i dan u_j	: koefisien pada variabel prediktor t
$k_1, k_2, \ldots, k_s$	: titik knot

sehingga model regresi *spline cubic* (2.27) apabila dinyatakan dalam bentuk matriks, maka diperoleh:

$$f\left(t\right) = z\alpha \tag{2.28}$$

$$\begin{bmatrix} f\left(t_1\right) \\ \vdots \\ f\left(t_n\right) \end{bmatrix} = \begin{bmatrix} 1 & t_1 & t_1^2 & t_1^3 & \left(t_1 - k_1\right)_+^3 & \left(t_1 - k_2\right)_+^3 & \cdots & \left(t_1 - k_s\right)_+^3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & t_n & t_n^2 & t_n^3 & \left(t_n - k_1\right)_+^3 & \left(t_n - k_2\right)_+^3 & \cdots & \left(t_n - k_s\right)_+^3 \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_{31} \\ \alpha_{32} \\ \vdots \\ \alpha_{3s} \end{bmatrix}$$

Sehingga model *spline cubic* pada regresi semiparametrik (2.5) dapat dinyatakan menjadi:

$$f\left(t_i\right) = \beta_0 + \beta_1 x_1 + \alpha_0 + \alpha_1 t_i + \alpha_2 t_i^2 + \alpha_3 t_i^3 + \alpha_{31}\left(t_i - k_1\right)_+^3 + \alpha_{32}\left(t_i - k_2\right)_+^3 + \ldots + \alpha_{3s}\left(t_i - k_s\right)_+^3 \tag{2.29}$$

dimana $i = 1, 2, \ldots, n$.

Apabila dinyatakan dalam bentuk matriks, maka diperoleh:

$$Y = X\beta + z\alpha + \varepsilon \tag{2.30}$$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} 1 & t_1 & t_1^{\,2} & t_1^{\,3} & (t_1-k_1)_+^3 & (t_1-k_2)_+^3 & \cdots & (t_1-k_s)_+^3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & t_n & t_n^{\,2} & t_n^{\,3} & (t_n-k_1)_+^3 & (t_n-k_2)_+^3 & \cdots & (t_n-k_s)_+^3 \end{bmatrix}$$

$$\begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_{31} \\ \alpha_{32} \\ \vdots \\ \alpha_{3s} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

2.13 Estimator Fungsi *Smoothing Spline* dalam Regresi Semiparametrik

Fungsi regresi nonparametrik pada persamaan (2.3), dimana $f(t_i)$ merupakan fungsi pemulus yang tidak spesifik, dengan ε_i adalah faktor pengganggu. Jika diamati variabel responnya y dan variabel bebas t, maka akan diperoleh pasangan pengamatan (t_i, y_i), $i = 1, 2, \dots, n$. $f(t_i)$ adalah kurva regresi yang belum diketahui dan $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T$ merupakan vektor *error* random independen (bebas) dengan mean 0 dan variansi σ^2, tetapi $f(t_i)$ hanya diasumsikan mulus (*smooth*). Fungsi *smooth* secara geometris, gradiennya berubah lambat sehingga dapat menggunakan suatu titik di sekitar titik tersebut sebagai investasinya. Diberikan $f(t_i)$ sebagai fungsi *smooth* yang termuat dalam suatu ruang fungsi tertentu, khususnya ruang Sobolev atau $f \in W_2^v[a, b]$, dengan

$$W_2^v[a,b] = \left\{ f; \left(\int_a^b f^{(v)}(t) \right)^2 dt < \infty \right\} \qquad (2.31)$$

dimana v bilangan positif untuk menyelesaikan estimasi kurva regresi, dan ε_i sesatan *random* yang diasumsikan berdistribusi normal dengan *mean* nol dan variansi σ^2 (Wahba, 1990).

Penalized kuadrat turunan ke-v memastikan *natural smoothing spline* pada orde ganjil $m=2v-1$ atau memiliki turunan ke-v atau ke-$(m-1)$ yang kontinu. Sehingga turunan fungsi $f(t)$ pada *smoothing spline cubic* dengan $m=3$ adalah:

$$Min_{f\in W_2^v[a,b]} \sum_{i=1}^{n}\left(y_i - x_i\beta - f(t_i)\right)^2 = Min_{f\in W_2^v[a,b]} \left(Y - X\beta - f\right)^2$$

$$= Min_{f\in W_2^v[a,b]} \left(Y - X\beta - f\right)^T \left(Y - X\beta - f\right) \quad (2.32)$$

dengan syarat, $g(f)=\int_a^b \left[f^{(v)}(t)\right]^2 dt \le \rho$, $\rho \ge 0$, sehingga taksiran ini ekuivalen dengan *Penalized Least Square* (*PLS*) pada regresi semiparametrik *smoothing spline*, yaitu penyelesaian optimasi:

$$PLS = \frac{1}{n}\sum_{i=1}^{n}\left(Y - x_i\beta - f(t)\right)^2 + \lambda\int_a^b \left(f^{(v)}(t)\right)^2 dt \quad (2.33)$$

Dari persamaan (2.33), $\sum_{i=1}^{n}\left(Y - x_i\beta - f(t)\right)^2$ merupakan *The Residual Sum of Square* (*RSS*) atau jumlah kuadrat sisaan, yang merupakan sebuah fungsi jarak antara data dan taksiran. Sedangkan $\lambda\int_a^b \left(f^{(v)}(t)\right)^2 dt$ merupakan *Penalized Roughness of The Function* yang diboboti dengan λ (parameter penghalus) yang memberikan ukuran kemulusan atau kekasaran kurva dalam memetakan data, melalui parameter penghalus $\lambda \geq 0$ (Thomas, 2003).

Penalized untuk *smoothing spline cubic* adalah kuadrat turunan ke-2, sehingga *smoothing spline cubic* pada model regresi semiparametrik menaksir λ untuk f didefinisikan sebagai optimasi *PLS*:

$$PLS = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - x_i\beta - f\left(t_i\right)\right)^2 + \lambda\int_a^b \left(f''\left(t_i\right)\right)^2 dt \qquad (2.34)$$

Dari persamaan (2.34), λ merupakan parameter penghalus dan penalti diberikan $\int_a^b \left(f''\left(t_i\right)\right)^2 dt$. Jika $\lambda \to 0$, maka hasil estimasi mendekati hasil metode kuadrat terkecil. Sebaliknya jika $\lambda \to \infty$, maka estimasi akan menginterpolasi titik-titik data. Estimator terbaik merupakan kompromi antara nilai jumlah kuadrat residual dan parameter penghalus λ yang bisa didapatkan dengan meminimumkan nilai GCV.

Dengan data amatan n sesuai dengan algoritma numerik dari persamaan matriks, maka persamaan (2.33) dapat ditulis

$$PLS = \left(Y - X\beta - f\right)^T\left(Y - X\beta - f\right) + \lambda f^T K f \qquad (2.35)$$

dimana K adalah matriks penalti yang memiliki struktur spesifik, yaitu: $K = QR^{-1}Q^T$ dimana Q adalah matriks $Q = q_{ij}$ yang berukuran $n \times \left(n-2\right)$ dan R adalah matriks $R = r_{ij}$ yang berukuran $\left(n-2\right)\times\left(n-2\right)$ (Wu dan Zhang, 2006).

2.14 Pemilihan Parameter Penghalus λ Optimal

Pada persamaan (2.34), pemilihan parameter penghalus λ yang optimal sangat penting untuk mendapatkan model estimator kurva yang baik. Parameter λ merupakan pengontrol keseimbangan antara kesesuaian kurva terhadap data dan kemulusan kurva. Pada nilai λ yang kecil maka kurvanya akan kasar atau sebaliknya, untuk nilai λ yang besar maka kurvanya akan menjadi mulus (*smooth*), dimana fungsi yang mulus terlihat jelas secara geometrik, ketika gradien kurva pada titik-titik knot tertentu tidak berubah dengan cepat (Eubank, 1999).

Kriteria yang biasa digunakan dalam pemilihan model *spline* terbaik adalah *Generalized Cross Validation* (GCV). Nilai GCV dipakai karena aspek perhitungannya lebih sederhana dan cukup efisien. Selain itu, kriteria model regresi yang umumnya dipakai masih tetap dijadikan acuan pemilihan model *spline* terbaik.

Model *spline* terbaik adalah yang mampu menjelaskan hubungan antara variabel prediktor dengan variabel respon dan memenuhi beberapa kriteria tertentu, antara lain:

a. Mempunyai nilai *Mean Square Error* (*MSE*) minimum

b. Menghasilkan nilai koefisien determinasi (R^2) maksimum

Nilai *MSE* merupakan nilai estimasi dari variansi residual sehingga model terbaik adalah model dengan *MSE* minimum yang menandakan nilai estimasi mendekati nilai sebenarnya. Sedangkan R^2 adalah alat untuk mengukur proporsi keragaman atau variansi total di sekitar nilai tengah y yang dapat dijelaskan oleh model regresi. Secara umum semakin besar nilai R^2, maka semakin baik pula model yang didapatkan karena mampu menjelaskan lebih banyak data (Draper dan Smith, 1998).

Nilai λ yang optimal berkaitan dengan nilai GCV_λ yang minimum. Pada model regresi *smoothing spline*, kriteria GCV didefinisikan sebagai:

$$GCV_\lambda = \frac{MSE(\lambda)}{\left[n^{-1}trace\left(I - A(\lambda)\right)\right]^2} \tag{2.36}$$

dimana $MSE(\lambda) = n^{-1}Y^T\left(I - A(\lambda)\right)^T\left(I - A(\lambda)\right)Y$ dan $Y = [Y_1, Y_1, Y_1, \ldots, Y_n]$

dengan

n : banyaknya data

GCV_λ : parameter penghalus

$MSE(\lambda)$: *mean square error* (*MSE*)

$A(\lambda)$: *hat matrix* berukuran $n \times n$

$trace\ A(\lambda)$: jumlah diagonal utama *hat matrix* perluasan dari parameter penghalus λ (Budiantara, 2000).

2.15 Menguji Signifikasi Model Regresi Semiparametrik

Pengujian parameter dalam model regresi bertujuan untuk mengetahui apakah parameter tersebut telah menunjukkan hubungan yang nyata antara variabel prediktor dan variabel respon. Selain itu juga untuk mengetahui kelayakan parameter dalam menerangkan model (Radythia, 2013).

2.15.1 Uji Simultan (Uji *F*)

Uji simultan digunakan untuk memeriksa signifikansi koefisien regresi secara bersama-sama (Supranto, 2005).

a. Hipotesis pengujian

$H_0 : \beta_p = 0,$ setiap $p = 0,1,2,\ldots,n$

lawan

H_1 : minimal terdapat satu $\beta_p \neq 0$, dengan $p = 0,1,2,\ldots,n$

b. Statistik uji

$$F = \frac{KT_{Reg}}{S^2} \tag{2.37}$$

c. Keputusan

Jika $F_{hitung} > F_{tabel}$ maka tolak H_0 dan terima H_1 (Draper dan Smith, 2003).

2.15.2 Uji Individu atau Uji Parsial (Uji *t*)

a. Hipotesis

$H_0 : \beta_p = 0,$ setiap $p = 0,1,2,\ldots,n$

$H_1 : \beta_p \neq 0,$ setiap $p = 0,1,2,\ldots,n$

b. Statistik uji

$$t = \frac{\hat{\beta}_j}{stdev\left(\hat{\beta}_p\right)} = \frac{\hat{\beta}_p}{\sqrt{\hat{\sigma}^2 C_p}} \tag{3.8}$$

C_p adalah elemen diagonal ke-i dari $(D'D)^{-1}$

c. Keputusan

Tolak H_0 jika $\left|t_{hitung}\right| > t_{\frac{\alpha}{2:df}}$ dengan $df = n - m$ merupakan banyaknya parameter (Radythia, 2013).

2.16 Kemiskinan

Definisi tentang kemiskinan telah mengalami perluasan, seiring dengan semakin kompleksnya faktor penyebab, indikator maupun permasalahan lain yang melingkupinya. Kemiskinan tidak lagi hanya dianggap sebagai dimensi ekonomi melainkan telah meluas hingga ke dimensi sosial, kesehatan, pendidikan, dan politik. Menurut Badan Pusat Statistik (BPS), kemiskinan adalah ketidakmampuan memenuhi standar minimum kebutuhan dasar yang meliputi kebutuhan makan maupun non makan.

Membandingkan tingkat konsumsi penduduk dengan garis kemiskinan atau jumlah rupiah untuk konsumsi orang perbulan. Definisi menurut *United Nations Development Programme* (UNDP) dalam Cahyat (2004), adalah ketidakmampuan untuk memperluas pilihan-pilihan hidup, antara lain dengan memasukkan penilaian tidak adanya partisipasi dalam pengambilan kebijakan publik sebagai salah satu indikator kemiskinan. Pada dasarnya definisi kemiskinan dapat dilihat dari dua sisi, yaitu:

a) Kemiskinan absolut

Kemiskinan yang dikaitkan dengan perkiraan tingkat pendapatan dan kebutuhan yang hanya dibatasi pada kebutuhan pokok atau kebutuhan dasar minimum yang memungkinkan seseorang untuk hidup secara layak. Dengan demikian kemiskinan diukur dengan membandingkan tingkat pendapatan orang dengan tingkat pendapatan yang dibutuhkan untuk memperoleh kebutuhan dasarnya yakni makanan, pakaian, dan perumahan agar dapat menjamin kelangsungan hidupnya.

b) Kemiskinan relatif

Kemiskinan dilihat dari aspek ketimpangan sosial, karena ada orang yang sudah dapat memenuhi kebutuhan dasar minimumnya tetapi masih jauh lebih rendah dibanding masyarakat sekitarnya (lingkungannya). Semakin besar ketimpangan antara tingkat penghidupan golongan atas dan golongan bawah maka akan semakin besar pula jumlah penduduk yang dapat dikategorikan miskin, sehingga kemiskinan relatif erat hubungannya dengan masalah distribusi pendapatan.

2.17 Tingkat Pengangguran Terbuka

Pengangguran adalah bagian dari angkatan kerja yang sekarang ini tidak bekerja dan sedang aktif mencari pekerjaan. Konsep ini sering diartikan sebagai keadaan pengangguran terbuka, yaitu orang yang tidak mempunyai pekerjaan, lengkapnya orang yang tidak bekerja dan (masih atau sedang) mencari pekerjaan. Masalah yang sering dihadapi adalah masalah setengah menganggur atau pengangguran tidak kentara.

- Setengah menganggur

Keadaan setengah menganggur (*underemployment*) terletak antara *full employment* dan sama sekali menganggur. Pengertian yang digunakan

International Labour Organization (ILO), *underemployment*, yaitu perbedaan antara jumlah pekerjaan yang betul dikerjakan seseorang dalam pekerjaannya dengan jumlah pekerjaan yang secara normal mampu dan ingin dikerjakannya. Konsep ini dibagi menjadi beberapa macam, yaitu:

a. Setengah menganggur yang kentara

Setengah menganggur yang kentara (*visible underemployment*) adalah jika seseorang bekerja tidak tetap (*part time*) di luar keinginannya sendiri, atau bekerja dalam waktu yang lebih pendek dari biasanya.

b. Setengah menganggur yang tidak kentara

Setengah menganggur yang tidak kentara (*invisible underemployment*) adalah jika seseorang bekerja secara penuh (*full time*) tetapi pekerjannya itu dianggap tidak mencukupi karena pendapatannya terlalu rendah atau pekerjaan tersebut tidak memungkinkan ia untuk mengembangkan seluruh keahliannya.

- Pengangguran tidak kentara

Pengangguran tidak kentara (*disguised unemployment*), dalam angkatan kerja mereka dimasukkan dalam kegiatan bekerja, tetapi sebetulnya mereka menganggur jika dilihat dari segi produktivitasnya. Jadi di sini mereka sebenarnya tidak mempunyai produktivitas dalam pekerjaannya. Misalnya mereka terdiri dari empat orang yang bersama-sama bekerja dalam jenis pekerjaan yang sesungguhnya dapat dikerjakan oleh tiga orang sehingga satu orang merupakan *disguised unemployment*.

- Pengangguran friksional

Pengangguran friksional, yaitu pengangguran yang terjadi akibat pindahnya seseorang dari suatu pekerjaan ke pekerjaan lain, dan akibatnya harus mempunyai waktu tenggang dan berstatus sebagai penganggur sebelum mendapatkan pekerjaan yang lain tersebut.

2.18 Indeks Pendidikan

Batasan pengertian pendidikan yang dikemukakan oleh para ahli tergantung dari sudut pandang yang dipergunakan dalam memberi arti pendidikan. Sudut pandang ini dapat bersumber dari aliran falsafah, pandangan hidup ataupun ilmu-ilmu pengetahuan yang berkaitan dengan tingkah laku manusia. Dalam UU RI No. 20 Tahun 2003 tentang sistem pendidikan nasional, pendidikan adalah usaha sadar dan terencana untuk mewujudkan suasana belajar dan proses pembelajaran agar peserta didik secara aktif mengembangkan potensi dirinya untuk memiliki kekuatan spiritual keagamaan, pengendalian diri, kepribadian, kecerdasan, akhlak mulia, serta keterampilan yang diperlukan dirinya, masyarakat, bangsa, dan negara.

Crow and Crow mendefinisikan pendidikan adalah proses yang berisi berbagai macam kegiatan yang sesuai dengan kegiatan seseorang untuk kehidupan sosialnya dan membantunya meneruskan kebiasaan dan kebudayaan, serta kelembagaan sosial dari generasi ke generasi (Idris dan Jamal, 1995).

Sedangkan menurut Frederick J. Donald, *education is the sense used here, in a process or an activity which is directed at producing desirable changes in the* *behavior of human beings*. Artinya pendidikan yang dimaksudkan di sini adalah proses atau aktivitas yang mengarah pada perubahan perilaku manusia (Donal, 1959).

Demikian beberapa pendapat tentang pendidikan, dari beberapa definisi di atas dapat penulis simpulkan bahwa pendidikan adalah:

1. Suatu pengarahan atau bimbingan yang diberikan kepada anak dalam pertumbuhannya
2. Suatu usaha sadar untuk menciptakan suatu keadaan atau situasi tentang yang dikehendaki oleh masyarakat
3. Suatu pembentukan kepribadian dan kemampuan anak menuju kedewasaan
4. Suatu bimbingan yang berperan untuk membentuk insan kamil

2.19 Inspirasi Al-Qur'an dalam Kajian tentang Estimasi

Adapun salah satu hadist, yaitu hadits Qudsi yang menginspirasi tentang estimasi dikatakan bahwa:

عن أبي هريرة – رضي الله عنه – قال,قال رسول الله– صلي الله عليه وسلم –: (إن الله يقول : أنا عند ظن عبدي بي, وآنامعه إذادعا ني)

Artinya:
Dari Abu Hurairah r.a, yang berkata bahwa Rasulullah SAW bersabda: "Sesungguhnya Allah SWT berfirman 'Aku menurut sangkaan hamba-Ku terhadap-Ku, dan aku bersamanya jika dia berdoa (minta tolong) kepada-Ku'". (HR. At-Tirmidzi, Husnu Adz-Dzaan billah. Hadist ini dikatakan hadist hasan shahih).

Penjelasan firman Allah SWT berbunyi "*Aku menurut sangkaan hamba-Ku terhadap-Ku*" adalah jika seorang hamba menyangka Allah SWT menerima amal shalih yang dilakukan maka Allah SWT akan menerima dan memberinya pahala yang setimpal, dan Allah SWT mengampuni jika seorang hamba tersebut benar-benar bertaubat. Demikian sebaliknya jika seorang hamba menyangka Allah SWT tidak akan menerima amal shalih yang dilakukan maka Allah SWT akan bertindak demikian.

Dalam hadits Qudsi tersebut jelas sekali bahwa apa yang disangkakan kepada Allah itu akan sama dengan apa yang akan Allah lakukan kepada hamba-Nya, dalam arti sangkaan hamba. Jika seorang hamba menyangka Allah akan mengampuni dosa-dosa hamba maka Allah akan mengampuninya, begitu pula sebaliknya, jika hamba menyangka bahwa Allah tidak akan mengampuni dosa hamba-Nya, maka Allah akan bertindak demikian. Hal ini dimaksudkan bahwa apa yang diestimasi oleh suatu estimator itu harus sama dengan apa yang diestimasi.

BAB III

PEMBAHASAN

3.1 Optimasi *Penalized Least Square* (*PLS*)

Untuk mendapatkan estimasi regresi semiparametrik *smoothing spline cubic* dilakukan dengan memaksimumkan fungsi *PLS* pada persamaan (2.35) dengan membuat $\frac{\partial PLS}{\partial \beta}=0$ dan $\frac{\partial PLS}{\partial f}=0$ sehingga diperoleh $\hat{\beta}$ dan $\hat{f}$.

$$PLS=\left(Y-X\beta-f\right)^T\left(Y-X\beta-f\right)+\lambda f^TKf$$

$$PLS=\left(Y^T-\beta^TX^T-f^T\right)\left(Y-X\beta-f\right)+\lambda f^TKf$$

$$PLS=Y^TY-Y^TX\beta-Y^Tf-\beta^TX^TY+\beta^TX^TX\beta+\beta^TX^Tf-f^TY+f^TX\beta+ f^Tf+\lambda f^TKf \tag{3.1}$$

$$PLS=Y^TY-2\beta^TX^TY-2f^TY+\beta^TX^TX\beta+\beta^TX^Tf+f^TX\beta+f^Tf+\lambda f^TKf \tag{3.2}$$

3.2 Estimasi Parameter β

Untuk mendapatkan estimasi parameter β, makan persamaan (3.2) diturunkan terhadap β dengan $\frac{\partial PLS}{\partial \beta}=0$ akan diperoleh:

$$\frac{\partial PLS}{\partial \beta}=0$$

$$-2X^TY+X^TX\beta+\left(\beta^TX^TX\right)^T+X^Tf+\left(f^TX\right)^T=0$$

$$-2X^TY+2X^TX\beta+2X^Tf=0$$

$$-X^TY+X^TX\beta+X^Tf=0 \qquad (3.3)$$

3.3 Estimasi Parameter *f*

Untuk memperoleh estimasi parameter f, maka persamaan (3.2) diturunkan terhadap f dengan $\frac{\partial PLS}{\partial f}=0$, diperoleh:

$$\frac{\partial PLS}{\partial f}=0$$

$$-2Y+\left(\beta^TX^T\right)^T+X\beta+f+\left(f^T\right)^T+2\lambda Kf=0$$

$$-2Y+2X\beta+2f+2\lambda Kf=0$$

$$-Y+X\beta+f+\lambda Kf=0$$

$$f+\lambda K^Tf=Y-X\beta$$

$$\left(I+\lambda K^T\right)f=Y-X\beta$$

$$f=\left(I+\lambda K^T\right)^{-1}\left(Y-X\beta\right) \qquad (3.4)$$

3.4 Menentukan $\hat{\beta}_\lambda$ dan $\hat{f}_\lambda$

Untuk memperoleh estimator $\hat{\beta}_\lambda$ dan $\hat{f}_\lambda$, yaitu dengan mensubstitusikan persamaan (3.3) ke persamaan (3.4):

$$-X^TY+X^TX\beta+X^Tf=0$$

$$-X^TY+X^TX\beta+X^T\left[\left(I+\lambda K^T\right)^{-1}\left(Y-X\beta\right)\right]=0$$

$$-X^TY + X^TX\beta + X^T\left(I + \lambda K^T\right)^{-1}\left(Y - X\beta\right) = 0$$

$$-X^TY + X^TX\beta + X^T\left(I + \lambda K^T\right)^{-1}Y - X^T\left(I + \lambda K^T\right)^{-1}X\beta = 0$$

$$X^TX\beta - X^T\left(I + \lambda K^T\right)^{-1}X\beta = X^TY - X^T\left(I + \lambda K^T\right)^{-1}Y$$

$$\left(X^TX - X^T\left(I + \lambda K^T\right)^{-1}X\right)\beta = \left(X^T - X^T\left(I + \lambda K^T\right)^{-1}\right)Y$$

$$\hat{\beta}_\lambda = \left(X^TX - X^T\left(I + \lambda K^T\right)^{-1}X\right)^{-1}\left(X^T - X^T\left(I + \lambda K^T\right)^{-1}\right)Y$$

$$\hat{\beta}_\lambda = \left(\left[X^T - X^T\left(I + \lambda K^T\right)^{-1}\right]X\right)^{-1}\left(X^T - X^T\left(I + \lambda K^T\right)^{-1}\right)Y$$

$$\hat{\beta}_\lambda = \left(X^T\left[I - \left(I + \lambda K^T\right)^{-1}\right]X\right)^{-1}\left(X^T\left[I - \left(I + \lambda K^T\right)^{-1}\right]\right)Y$$

$$\hat{\beta}_\lambda = \left(X^TPX\right)^{-1}\left(X^TP\right)Y \qquad (3.5)$$

dimana $P = I - \left(I + \lambda K^T\right)^{-1}$

Substitusikan persamaan (3.5) ke dalam persamaan (3.4) sehingga akan diperoleh:

$$f = \left(I + \lambda K^T\right)^{-1}\left(Y - X\beta\right)$$

$$f = \left(I + \lambda K^T\right)^{-1}\left(Y - X\left[\left(X^TPX\right)^{-1}\left(X^TP\right)Y\right]\right)$$

$$\hat{f}_\lambda = \left(I + \lambda K^T\right)^{-1}\left(Y - X\left(X^TPX\right)^{-1}\left(X^TP\right)Y\right)$$

$$\hat{f}_\lambda = \left(I + \lambda K^T\right)^{-1}\left[I - X\left(X^TPX\right)^{-1}\left(X^TP\right)\right]Y \qquad (3.6)$$

3.5 Menentukan *Hat Matrix* $A(\lambda)$

Untuk mendapatkan matriks $A(\lambda)$, substitusikan persamaan (3.5) dan persamaan (3.6) ke

$$\hat{Y} = X\hat{\beta}_\lambda + \hat{f}_\lambda$$

$$A(\lambda)Y = X\left[\left(X^TPX\right)^{-1}\left(X^TP\right)Y\right] + \left(I+\lambda K^T\right)^{-1}\left[I - X\left(X^TPX\right)^{-1}\left(X^TP\right)\right]Y$$

$$A(\lambda)Y = X\left(X^TPX\right)^{-1}\left(X^TP\right)Y + \left[\left(I+\lambda K^T\right)^{-1} - \left(I+\lambda K^T\right)^{-1}X\left(X^TPX\right)^{-1}\left(X^T\mathbf{P}\right)\right]\mathbf{Y}$$

$$A(\lambda)Y = \left[X\left(X^TPX\right)^{-1}\left(X^TP\right) + \left(I+\lambda K^T\right)^{-1} - \left(I+\lambda K^T\right)^{-1}X\left(X^TPX\right)^{-1}\left(X^TP\right)\right]Y$$

$$A(\lambda) = X\left(X^TPX\right)^{-1}\left(X^TP\right) - \left(I+\lambda K^T\right)^{-1}X\left(X^TPX\right)^{-1}\left(X^TP\right) + \left(I+\lambda K^T\right)^{-1}$$

$$A(\lambda) = \left[I - \left(I+\lambda K^T\right)^{-1}\right]X\left(X^TPX\right)^{-1}\left(X^TP\right) + \left(I+\lambda K^T\right)^{-1}$$

$$A(\lambda) = PX\left(X^TPX\right)^{-1}\left(X^TP\right) + \left(I+\lambda K^T\right)^{-1} \qquad (3.7)$$

3.6 Menentukan Parameter Penghalus Optimal dengan Metode GCV

Untuk menentukan parameter penghalus λ yang optimal dengan menggunakan metode GCV_λ. Fungsi GCV_λ (2.36) didefinisikan sebagai berikut:

$$GCV_\lambda = \frac{MSE(\lambda)}{\left[n^{-1}trace\left(I - A(\lambda)\right)\right]^2}$$

$$GCV_\lambda = \frac{n^{-1}Y^T\left[\left(I - A(\lambda)\right)^T\left(I - A(\lambda)\right)\right]Y}{\left[n^{-1}trace\left(I - A(\lambda)\right)\right]^2} \qquad (3.8)$$

dimana

$Y = \left(Y_{11}, Y_{12}, Y_{13}, \ldots, Y_{1n}, \ldots, Y_{n1}, Y_{n2}, \ldots, Y_{nm}\right)$ dan

$A(\lambda) = PX\left(X^TPX\right)^{-1}\left(X^TP\right) + \left(I+\lambda K^T\right)^{-1}$.

Selanjutnya untuk pemilihan parameter panghalus λ optimal dengan menggunakan metode GCV_λ, dilakukan dengan cara mensubstitusikan nilai

$0 < \lambda < 1$ ke dalam matriks $A(\lambda)$ pada persamaan (3.8) hingga diperoleh nilai minimum GCV_{λ}.

3.7 Regresi Semiparametrik untuk Memodelkan Kemiskinan di Kota Malang Tahun 2005-2012

3.7.1 Deskripsi Data

Data yang dipakai dalam penelitian ini merupakan data kemiskinan di Kota Malang pada tahun 2005-2012 yang bersumber dari Badan Pusat Statistik (BPS) Kota Malang. Kemiskinan ini dipengaruhi oleh tingkat pengangguran terbuka sebagai variabel prediktor unsur parametrik dan indeks pendidikan sebagai variabel prediktor unsur nonparametrik dengan melakukan pengujian. Sehingga untuk variabel-variabel dalam regresi semiparametrik untuk memodelkan kemiskinan di Kota Malang tahun 2005-2012 adalah sebagai berikut:

Y : jumlah kemiskinan di Kota Malang tahun 2005-2012

X : tingkat pengangguran terbuka di Kota Malang tahun 2005-2012

t : indeks pendidikan di Kota Malang tahun 2005-2012.

3.7.2 Mendeteksi Variabel Prediktor Komponen Parametrik dan Variabel Nonparametrik Menggunakan *Scatter Plot*

Untuk mengetahui data tersebut dalam kelompok parametrik atau nonparametrik, terlebih dahulu akan dilakukan uji normalitas terhadap data dengan cara melihat *plot* dari data tersebut dengan menggunakan hipotesis sebagai berikut:

$H_0 : \mu = 0$ (normal)

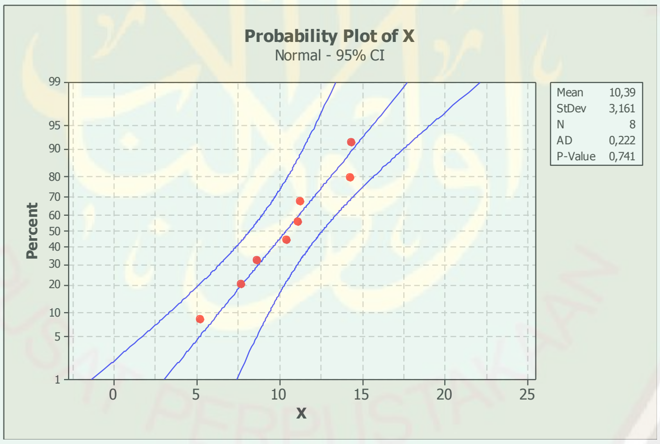
$H_1 : \mu \neq 0$ (tidak normal)

Dalam hal pengujian hipotesis ini, kriteria untuk menolak atau menerima H_0 berdasarkan $p-value$ atau nilai signifikansi uji yang dinyatakan sebagai berikut:

Jika $p-value < \alpha = 0.05$, maka H_0 ditolak

Jika $p-value > \alpha = 0.05$, maka H_0 diterima

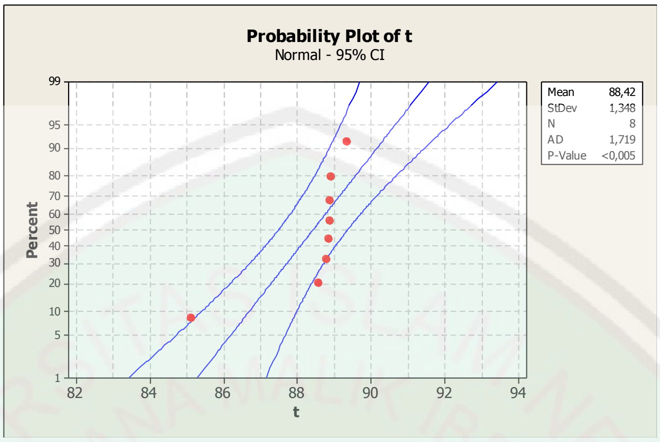
Dengan menggunakan program MINITAB 14, *scatter plot* untuk uji normalitas pada tingkat pengangguran terbuka adalah sebagai berikut:



Gambar 3.1: *Plot* untuk Uji Normalitas pada Tingkat Pengangguran Terbuka

Gambar di atas menunjukkan bahwa $p-value > 0.05$, maka terima H_0 dan tolak H_1 yang berarti bahwa data menyebar normal. Sehingga data tingkat pengangguran terbuka terhadap jumlah kemiskinan merupakan data untuk regresi parametrik.

Untuk uji normalitas pada indeks pendidikan adalah sebagai berikut:



Gambar 3.2: *Plot* untuk Uji Normalitas pada Indeks Pendidikan

Gambar di atas menunjukkan bahwa $p - value < 0.05$, maka tolak H_0 dan terima H_1 yang berarti bahwa data tidak menyebar normal. Sehingga data indeks pendidikan terhadap jumlah kemiskinan merupakan data untuk regresi nonparametrik.

Sehingga dapat diambil kesimpulan untuk komponen parametrik adalah tingkat pengangguran terbuka dan untuk komponen nonparametrik adalah indeks pendidikan. Dari uji normalitas di atas data tingkat pengangguran terbuka dan indeks pendidikan digabung menjadi data untuk regresi semiparametrik karena telah memenuhi asumsi-asumsi regresi semiparametrik, yaitu sebagian bentuk kurva diketahui dan sebagian bentuk kurva belum diketahui.

Sehingga model yang digunakan untuk data tersebut adalah persamaan (2.29) sebagai berikut:

$$y_i = \beta_0 + \beta_1 + \alpha_0 + \alpha_1 t_i + \alpha_2 t_i^{\,2} + \alpha_3 t_i^{\,3} + \alpha_{31}(t_i - k_1)^3 + \alpha_{32}(t_i - k_2)^3 + \cdots + \alpha_{3s}(t_i - k_s)^3$$

dimana $i = 1, 2, ..., n$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} 1 & t_1 & t_1^{\,2} & t_1^{\,3} & (t_1 - k_1)_+^3 & (t_1 - k_2)_+^3 & \cdots & (t_1 - k_s)_+^3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_n & t_n^{\,2} & t_n^{\,p} & (t_n - k_1)_+^3 & (t_n - k_2)_+^3 & \cdots & (t_n - k_s)_+^3 \end{bmatrix}$$

$$\begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_{31} \\ \alpha_{32} \\ \vdots \\ \alpha_{3s} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

3.7.3 Memilih Model Tentatif Terbaik untuk S*moothing Spline Cubic*

Dalam membentuk suatu model regresi *spline cubic*, dibutuhkan informasi tentang banyaknya titik knot. Pada penelitian ini terdiri 8 data pengamatan, untuk itu terdapat 6 kombinasi titik knot yang ditentukan secara acak untuk membentuk model *spline cubic*. Dalam penelitian ini jumlah titik knot dibatasi sebanyak satu titik knot. Pada hakekatnya, pemilihan parameter penghalus sama dengan memilih titik knot. Parameter penghalus $0 < \lambda < 1$ sehingga dalam penelitian ini diambil satu parameter penghalus. Dengan bantuan MATLAB, maka diperoleh model tentative, nilai *MSE* dan GCV ditunjukkan pada tabel 3.1 sebagai berikut:

Tabel 3.1 Model Tentatif *Smoothing Spline Cubic* dengan semua kemungkinan satu parameter penghalus dan nilai *MSE* beserta GCVnya

λ	Model Tentatif	Nilai *MSE* Optimal	Nilai GCV Optimal
0.0100	$\hat{Y} = 3.4790 + 0.3643X + 0.8487 + 1.7993t + 1.7315t^2 + 0.6920t^3 + 1.7073(t - 0.01)^3$	1.4580	3.6900
0.1500	$\hat{Y} = 6.0594 - 0.3657X - 0.3999 - 1.2588t +$	0.9320	1.8343

	$0.2107t^2+2.0294t^3+0.2866(t-0.15)^3$		
0.2900	$\hat{Y}=5.9332-0.1379X-0.1738-1.0414t+0.4058t^2+2.1805t^3+0.4691(t-0.29)^3$	0.6919	1.5703
0.4300	$\hat{Y}=5.8997+0.0493X-0.0856-0.9556t+0.4858t^2+2.2488t^3+0.5457(t-0.43)^3$	0.6348	1.5028
0.5700	$\hat{Y}=5.8842-0.0001X-0.0366-0.9077t+0.5309t^2+2.2885t^3+0.5893(t-0.57)^3$	0.6093	1.4720
0.7100	$\hat{Y}=5.8753+0.0314X-0.0052-0.8769t+0.5601t^2+2.3146t^3+0.6177(t-0.71)^3$	0.5949	1.4544
0.8500	$\hat{Y}=5.8695+0.0534X+0.0167-0.8554t+0.5806t^2+2.3331t^3+0.6376(t-0.85)^3$	0.5856	1.4430
0.9900	$\hat{Y}=5.8654+0.0697X+0.0329-0.8395t+0.5958t^2+2.3469t^3+0.6524(t-0.99)^3$	**0.5792**	**1.4350**

(Sumber: Analisis Penulis dengan MATLAB)

Model regresi semiparametrik terbaik dipilih berdasarkan nilai dan minimum dari *MSE* dan GCV dengan nilai terkecil yang diperoleh dari semua kemungkinan kombinasi parameter penghalus dengan menggunakan pendekatan fungsi *smoothing spline cubic* yang disajikan pada tabel 3.1. nilai *MSE* dan GCV yang paling kecil pada $\lambda = 0.99$. Sehingga model tentatif regresi semiparametrik *smoothing spline cubic* dengan satu parameter penghalus sebagai berikut:

$$\hat{Y}=5.8654+0.0697X+0.0329-0.8395t+0.5958t^2+2.3469t^3+0.6524(t-0.99)^3 \quad (3.9)$$

3.8 Menguji Signifikansi Model Regresi Semiparametrik

3.8.1 Uji Simultan

Uji simultan merupakan uji hipotesis yang dipakai untuk mengetahui pengaruh secara bersama-sama dengan menggunakan tarif signifikansi 5% dan menggunakan hipotesis sebagai berikut:

$H_0 : \beta_j = 0$ setiap $'j = 0,1,2,3,4$

$H_1 :$ minimal terdapat satu $\beta_j \neq 0$; $j = 1,2,3,4$

Tabel 3.2 ANOVA Regresi *Smoothing Spline Cubic* Satu Parameter Penghalus

Sumber Variasi	db	JK	KT	Fhitung	F(*t*)
Regresi	2	194805558	97402779	3,89	0,096
Residual	5	125248642	25049728		
Total	7	320054200			

(Sumber: Analisis Penulis dengan MATLAB)

Dari tabel di atas diketahui bahwa $F_{hitung} > F_{tabel}$ sehingga menolak H_0 dan menerima H_1 yang berarti bahwa terdapat minimal satu parameter yang signifikan.

3.8.2 Uji Asumsi

a. Uji Autokolerasi

Uji autokorelasi merupakan salah satu pengujian yang dilakukan untuk mengetahui adanya korelasi di antara *error*. Dengan menggunakan persamaan Durbin Watson dengan hipotesis sebagai berikut:

$H_0 : \rho = 0$, tidak ada autokorelasi antara *error*

$H_1 : \rho \neq 0$, ada autokorelasi antara *error*

diperoleh,

$$DW = \frac{\sum_{t=1}^{5}\left(e_t - e_{t-1}\right)^2}{\sum_{t=1}^{5} e_t^{\,2}}$$

$$= 1.01415$$

Dari hasil perhitungan *DW* di atas diperoleh nilai $DW < 1.1$, sehingga tolak H_0 dan terima H_1 yang dapat disimpulkan adanya autokorelasi antar *error*. Berarti di antara *error* terdapat korelasinya. Sehingga regresi semiparametrik pada data jumlah kemiskinan di Kota Malang terpenuhi asumsi regresi semiparametriknya.

b. Uji Non-multikolinieritas

Uji ini dilakukan untuk mendeteksi apakah antara variabel bebas terjadi korelasi atau tidak. Uji adanya multikolinieritas ini dilihat dari *Variance Inflation Factor* (VIF). Dapat dilihat pada lampiran 2, bahwa nilai VIF tiap variabel bebas tidak lebih dari 10. Hal ini menunjukkan bahwa antara variabel bebas terjadi non-multikolinieritas. Berarti antara variabel bebas terjadi korelasi. Sehingga regresi semiparametrik pada data jumlah kemiskinan di Kota Malang terpenuhi asumsi regresi semiparametriknya.

3.8.3 Uji Kelayakan Model

Dalam hal ini uji kelayakan model menggunakan koefisien determinasi $\left(R^2\right)$ yang berguna untuk mengetahui besarnya pengaruh variabel respon terhadap variabel prediktor.

$$R^2 = \frac{\sum\left(\hat{y}_i - \bar{y}\right)}{\sum\left(y_i - \bar{y}\right)}$$

$$= 9.4\%$$

artinya faktor-faktor yang mempengaruhi jumlah kemiskinan di Kota Malang dijelaskan oleh tingkat pengangguran terbuka hanya sebesar 0.094 atau sebesar 9.4% sedangkan selebihnya 90.6% dijelaskan oleh faktor lain yang tidak terdapat pada model. Meskipun koefisien determinasi yang dihasilkan begitu kecil, tetapi paling tidak bisa digunakan untuk mengetahui faktor-faktor yang mempengaruhi jumlah kemiskinan di Kota Malang sehingga dapat menjadi perhatian untuk menurunkan jumlah kemiskinan di Kota Malang.

3.9 Inspirasi Al-Qur'an dalam Kajian tentang Estimasi Parameter

Adapun salah satu ayat Al-Qur'an yang menginspirasi tentang estimasi dalam surat Ash Shâffât (37) ayat 50 yang berbunyi:

فَأَقْبَلَ بَعْضُهُمْ عَلَى بَعْضٍ يَتَسَاءَلُونَ

Artinya: *"Lalu sebagian mereka menghadap kepada sebagian yang lain sambil bercakap-cakap"* (QS. Ash Shâffât: 50).

Dari ayat di atas, Allah SWT menerangkan bahwa orang-orang mukmin dalam surga duduk saling berhadap-hadapan dan berbincang-bincang satu sama lain sambil menikmati minuman yang disuguhkan kepada mereka. Betapa nikmatnya mengenang masa lampau mereka sewaktu dalam kesenangan dan ketenteraman hidup dalam surga itu. Mereka berbincang-bincang tentang berbagai keutamaan dan pengalaman mereka di dunia. Mereka saling memperbincangkan ahli surga dan neraka, keadaan orang-orang yang hidup berbahagia dan hidup sengsara. Mereka juga menengok kedua golongan ini dan ganjaran yang mereka peroleh dan hukuman yang mereka derita juga memperbincangkan percakapan keadaan orang-orang yang berada di antara surga dan neraka. (*Lalu sebagian* *mereka menghadap*), yakni sebagian penduduk surga (*kepada sebagian yang lain sambil bercakap-cakap*) mengenai apa yang telah mereka lakukan di dunia. Dari kata "*sebagian dari mereka"* merupakan suatu estimasi atau taksiran. Dimana kata *"Sebagian"* jumlahnya belum diketahui, oleh karena itu perlu diestimasi jumlahnya.

Dalam menduga berapa banyaknya jumlah kemiskinan yang harus diminimumkan harus memperhatikan beberapa faktor yang mungkin saja mempengaruhi tingkat kemiskinan di Kota Malang seperti halnya tingkat pengangguran terbuka dan indeks pendidikan. Setelah mengetahui faktor-faktor tersebut maka paling tidak hal ini dapat digunakan untuk memprediksi jumlah kemiskinan pada tahun berikutnya sehingga tingkat kemiskinan dapat diminimumkan dengan berbagai macam cara.

BAB IV

PENUTUP

1.1 Kesimpulan

Dari pembahasan di atas dapat disimpulkan bahwa:

1. Menentukan parameter penghalus optimal pada *smoothing spline* dengan metode GCV:

   a. Estimasi parameter dari regresi semiparametrik *smoothing spline cubic* $Y = X\beta + f + \varepsilon$ dengan menggunakan metode *Penalized Least Square* (*PLS*) adalah

   $$\hat{\beta}_\lambda = \left(X^T P X\right)^{-1}\left(X^T P\right)Y$$

   $$\hat{f}_\lambda = \left(I + \lambda K^T\right)^{-1}\left[I - X\left(X^T P X\right)^{-1}\left(X^T P\right)\right]Y$$

   b. *Hat matrix* dari hasil estimasi $\hat{\beta}_\lambda$ dan $\hat{f}_\lambda$ adalah

   $$A(\lambda) = PX\left(X^T P X\right)^{-1}\left(X^T P\right) + \left(I + \lambda K^T\right)^{-1}$$

   dengan $P = I - \left(I + \lambda K^T\right)^{-1}$.

   c. Fungsi $GCV_\lambda = \dfrac{n^{-1}Y^T\left[\left(I - A(\lambda)\right)^T\left(I - A(\lambda)\right)\right]Y}{\left[n^{-1}trace\left(I - A(\lambda)\right)\right]^2}$

   dan untuk memilih parameter panghalus λ optimal dengan menggunakan metode GCV_λ di atas, dilakukan dengan cara mensubstitusikan nilai $0<\lambda<1$ ke dalam *hat matrix* $A(\lambda)$ pada persamaan GCV_λ hingga diperoleh nilai minimum GCV_λ.

2. Analisis data kemiskinan di Kota Malang pada tahun 2005-2012, yaitu:

   a. Model regresi semiparametrik *smoothing spline cubic* dengan nilai GCV paling optimal pada $\lambda = 0.99$ adalah

   $$\hat{Y} = 5.8654 + 0.0697X + 0.0329 - 0.8395t + 0.5958t^2 + 2.3469t^3 + 0.6524(t-0.99)^3$$

   b. Setelah dilakukan uji asumsi normalitas bahwa data tingkat pengangguran terbuka terhadap jumlah kemiskinan berdistribusi normal dan data indeks pendidikan terhadap jumlah kemiskinan yang berdistribusi tidak normal, pada uji autokorelasi telah tejadi autokorelasi, uji non-multikolinieritas telah terjadi non-multikolinieritas dan uji kelayakan model dapat dilihat bahwa yang mempengaruhi kemiskinan di Kota Malang adalah tingkat pengangguran terbuka dan indeks pendidikan dengan $R^2 = 9.4\%$. Hal ini dapat disimpulkan bahwa data tingkat pengangguran terbuka dan indeks pendidikan terhadap jumlah kemiskinan telah memenuhi asumsi-asumsi regresi semiparametrik.

1.2 Saran

Diharapkan untuk penelitian selanjutnya menggunakan estimator yang lain untuk mencari estimasi parameternya dan juga bisa menggunakan fungsi yang lain selain fungsi *smoothing spline cubic* untuk regresi semiparametrik dan menggunakan data dengan banyak variabel prediktornya.

DAFTAR PUSTAKA

Aydin, D. dan Tuzemen, M.S.. 2012. Smoothing Parameter selection Problem in Nonparametric Regression Based on Smoothing Spline. *Journal of Applied Science*. 12: 636-644.

Algifari. 2000. *Analisis Regresi (Teori dan Kasus, edisi 2)*. Yogyakarta: BPFE Yogyakarta.

Animous. 2013. *Jarque-Berra Test*. http://www.wikipedia.com. Tanggal akses: 8 Sepetember 2013.

Budiantara, I.N.. 2000. Metode U, GML, CV dan GCV dalam Regresi Nonparametrik Spline. *Majalah Ilmiah Himpunan Matematika Indonesia (MIHMI)*, Vol. 6, 285– 290.

Budiantara, I.N.. 2005b. Penentuan Titik-Titik Knots dalam Regresi Spline. *Jurnal Jurusan Statistika FMIPA-ITS*. Suraba

Cahyat. 2004. Bagaimana Kemiskinan Diukur ? Beberapa Model Penghitungan Kemiskinan di Indonesia. *Governance Brief, 21 - 8.*

Donald, F.J.. 1959. *Education Psychology*. Tokyo: Wadsworth Publishing Company.

Draper, N.R. and Smith, H.. 1998, *Applied Regression Analysis*, Third Edition, New York: John Wiley and Sons.

Eubank, R.L.. 1999. *Nonparametric Regression and Smoothing Spline*. New York: Marcel Dekker INC.

Fan, J. dan Yao, Q.. 2005. *Nonlinear Time Series Nonparametric and Parametric Method*. Canada: Springer Science.

Firdaus, M.. 2004. *Ekonometrika Suatu Pendekatan Aplikatif*. Jakarta: Bumi Aksara.

Hasan, M.I.. 2001. *Pokok-pokok Materi Statistik I (Statistik Deskriptif)*. Jakarta: Bumi Aksara.

Idris, H.Z. dan Jamal, H.L.. 1995. *Pengantar Pendidikan*. Jakarta: Grasindo.

Kurniawan, D.. 2008. *Regresi Linier (Linear Regression)*: Forum Statistika.

Lee, T.C.M.. 2003. Smoothing Parameter Selection for Smoothing Spline: a simulation study. *Computational Statistics & data Analysis*. 42:139-148.

Marimba, A.D.. 1986. *Pengantar Filsafat Pendidikan Islam*. Bandung: PT. A-Ma'arif.

Naseem. 2010. *Fundamental Numerical Analysis and Error Estimation*. Anamaya Publisher, New Delhi.

Radythia, N.. 2013. *Pendekatan Regresi Semiparametrik untuk Proses Pembentukan Limbah Pabrik Gula Asembagus Situbondo*. http://digilib.its.ac.id/ITS-Undergraduate-13371-Paper.pdf tanggal akses: 09 sepember 2013.

Ruppert, W. dan Carrol. 2003. *Semiparametric regression*. United Kingdom: Chambrige University.

Supranto, J.. 2005. *Ekonometri (Buku Kesatu)*. Bogor: Ghalia Indonesia.

Wahba G.. 1990. *Splines Models for Observational Data*, SIAM, Philadelphia. CBMS-NSF Regional Conference Series in Applied Mathematics, Vol. 59.

Wahyu, W.. 2009. Metode Kuadrat Terkecil untuk Estimasi Kurva Regresi Semiparametrik Spline. *Seminar Nasional Matematika dan Pendidikan Matematika*. Vol 13, Hal 633-645.

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

Yatchew, A.. 2003. *Semiparametric Regression for the Applied Econometrian*. New York: Cambridge University Press.

Lampiran 1 Data Pengamatan

Y Jumlah Kemiskinan	X Tingkat Pengangguran Terbuka	t Indeks Pendidikan
54800	14,38	85,12
59400	14,31	88,58
58330	11,27	88,79
57200	11,14	88,87
44370	10,44	88,84
48400	8,68	88,87
45440	5,19	88,92
43100	7,68	89,33

Lampiran 2 Hasil Perhitungan *MSE* dengan MATLAB

```
MSE1 =  1.4580e+008
MSE2 =  9.3201e+007
MSE3 =  6.9192e+007
MSE4 =  6.3484e+007
MSE5 =  6.0935e+007
MSE6 =  5.9492e+007
MSE7 =  5.8563e+007
MSE8 =  5.7916e+007
MSE =
  1.0e+008 *
    1.4580
    0.9320
    0.6919
    0.6348
    0.6093
    0.5949
    0.5856
    0.5792
MSEmin=  5.7916e+007
```

Lampiran 3 Hasil Perhitungan GCV dengan MATLAB

```
GCV1 =  3.6900e+008
GCV2 =  1.8343e+008
GCV3 =  1.5703e+008
GCV4 =  1.5028e+008
GCV5 =  1.4720e+008
GCV6 =  1.4544e+008
GCV7 =  1.4430e+008
GCV8 =  1.4350e+008

GCV =
1.0e+008 *
    3.6900
    1.8343
    1.5703
    1.5028
    1.4720
    1.4544
    1.4430
    1.4350

GCVmin =  1.4350e+008
```

Lampiran 4 Hasil Perhitungan *Hat Matrix* dengan MATLAB

A1 =

0.4655	0.4283	0.2504	0.2612	-0.3848	-0.0295	0.0299	-0.0072
0.4283	0.3827	0.2189	0.1360	-0.1307	-0.0003	0.0146	-0.0019
0.2504	0.2189	0.0025	0.0537	0.3903	-0.0162	-0.0175	0.0088
0.2612	0.1360	0.0537	-1.5664	1.8569	0.3775	-0.1277	0.0368
-0.3848	-0.1307	0.3903	1.8569	-0.0533	-0.7171	-0.0736	-0.0100
-0.0295	-0.0003	-0.0162	0.3775	-0.7171	1.5184	0.0349	-0.1025
0.0299	0.0146	-0.0175	-0.1277	-0.0736	0.0349	1.1848	0.0634
-0.0072	-0.0019	0.0088	0.0368	-0.0100	-0.1025	0.0634	1.0373

A2 =

0.4980	0.4367	0.2107	0.0459	-0.0464	-0.0669	-0.1121	-0.0351
0.4367	0.3879	0.2009	0.0700	0.0000	-0.0223	-0.0490	-0.0153
0.2107	0.2009	0.1314	0.1160	0.1035	0.0873	0.1061	0.0263
0.0459	0.0700	0.1160	0.1515	0.2368	0.2012	0.2808	0.1052
-0.0464	0.0000	0.1035	0.2368	0.2364	0.2791	0.3749	0.0606
-0.0669	-0.0223	0.0873	0.2012	0.2791	0.0688	0.2634	0.2417
-0.1121	-0.0490	0.1061	0.2808	0.3749	0.2634	0.0068	-0.3982
-0.0351	-0.0153	0.0263	0.1052	0.0606	0.2417	-0.3982	0.8167

A3 =

0.5012	0.4380	0.2070	0.0372	-0.0535	-0.0677	-0.0930	-0.0280
0.4380	0.3886	0.1992	0.0664	-0.0033	-0.0227	-0.0404	-0.0121
0.2070	0.1992	0.1380	0.1222	0.1109	0.0857	0.0899	0.0206
0.0372	0.0664	0.1222	0.1854	0.2488	0.2046	0.2285	0.0850
-0.0535	-0.0033	0.1109	0.2488	0.2840	0.2670	0.3147	0.0417
-0.0677	-0.0227	0.0857	0.2046	0.2670	0.1522	0.2014	0.2059
-0.0930	-0.0404	0.0899	0.2285	0.3147	0.2014	0.1903	-0.3223
-0.0280	-0.0121	0.0206	0.0850	0.0417	0.2059	-0.3223	0.8499

A4 =

0.5021	0.4384	0.2059	0.0347	-0.0554	-0.0676	-0.0879	-0.0262
0.4384	0.3888	0.1986	0.0653	-0.0043	-0.0227	-0.0380	-0.0112
0.2059	0.1986	0.1401	0.1239	0.1129	0.0848	0.0855	0.0192
0.0347	0.0653	0.1239	0.1958	0.2516	0.2046	0.2145	0.0796
-0.0554	-0.0043	0.1129	0.2516	0.2984	0.2622	0.2984	0.0369
-0.0676	-0.0227	0.0848	0.2046	0.2622	0.1771	0.1852	0.1960
-0.0879	-0.0380	0.0855	0.2145	0.2984	0.1852	0.2392	-0.3021
-0.0262	-0.0112	0.0192	0.0796	0.0369	0.1960	-0.3021	0.8587

A5 =

0.5026	0.4386	0.2054	0.0335	-0.0563	-0.0675	-0.0855	-0.0253
0.4386	0.3889	0.1984	0.0648	-0.0047	-0.0227	-0.0369	-0.0108
0.2054	0.1984	0.1412	0.1248	0.1139	0.0844	0.0835	0.0185
0.0335	0.0648	0.1248	0.2009	0.2528	0.2045	0.2081	0.0772
-0.0563	-0.0047	0.1139	0.2528	0.3054	0.2597	0.2908	0.0347
-0.0675	-0.0227	0.0844	0.2045	0.2597	0.1892	0.1778	0.1914
-0.0855	-0.0369	0.0835	0.2081	0.2908	0.1778	0.2618	-0.2928
-0.0253	-0.0108	0.0185	0.0772	0.0347	0.1914	-0.2928	0.8628

A6 =

0.5029	0.4387	0.2050	0.0328	-0.0568	-0.0674	-0.0841	-0.0248
0.4387	0.3890	0.1982	0.0645	-0.0050	-0.0227	-0.0363	-0.0106
0.2050	0.1982	0.1418	0.1253	0.1144	0.0841	0.0823	0.0181
0.0328	0.0645	0.1253	0.2039	0.2534	0.2044	0.2043	0.0757
-0.0568	-0.0050	0.1144	0.2534	0.3095	0.2582	0.2863	0.0335
-0.0674	-0.0227	0.0841	0.2044	0.2582	0.1963	0.1736	0.1887
-0.0841	-0.0363	0.0823	0.2043	0.2863	0.1736	0.2749	-0.2874
-0.0248	-0.0106	0.0181	0.0757	0.0335	0.1887	-0.2874	0.8652

A7 =

0.5030	0.4388	0.2048	0.0323	-0.0572	-0.0674	-0.0832	-0.0245
0.4388	0.3890	0.1981	0.0644	-0.0051	-0.0226	-0.0359	-0.0105
0.2048	0.1981	0.1423	0.1256	0.1148	0.0839	0.0815	0.0179
0.0323	0.0644	0.1256	0.2059	0.2538	0.2043	0.2019	0.0748
-0.0572	-0.0051	0.1148	0.2538	0.3122	0.2571	0.2835	0.0327
-0.0674	-0.0226	0.0839	0.2043	0.2571	0.2009	0.1708	0.1870
-0.0832	-0.0359	0.0815	0.2019	0.2835	0.1708	0.2834	-0.2839
-0.0245	-0.0105	0.0179	0.0748	0.0327	0.1870	-0.2839	0.8667

A8 =

0.5032	0.4389	0.2047	0.0320	-0.0574	-0.0673	-0.0826	-0.0243
0.4389	0.3890	0.1981	0.0642	-0.0052	-0.0226	-0.0356	-0.0104
0.2047	0.1981	0.1426	0.1258	0.1150	0.0838	0.0810	0.0177
0.0320	0.0642	0.1258	0.2073	0.2541	0.2042	0.2001	0.0742
-0.0574	-0.0052	0.1150	0.2541	0.3142	0.2564	0.2815	0.0321
-0.0673	-0.0226	0.0838	0.2042	0.2564	0.2042	0.1689	0.1857
-0.0826	-0.0356	0.0810	0.2001	0.2815	0.1689	0.2894	-0.2814
-0.0243	-0.0104	0.0177	0.0742	0.0321	0.1857	-0.2814	0.8678

Lampiran 5 ***Output* MINITAB**

Regression Analysis: Y versus X; t

```
The regression equation is
Y = - 66583 + 1906 X + 1110 t


Predictor     Coef  SE Coef      T      P  VIF
Constant    -66583   158905  -0,42  0,693
X           1905,8    743,5   2,56  0,050  1,5
t             1110     1744   0,64  0,552  1,5


S = 5004,97   R-Sq = 60,9%   R-Sq(adj) = 45,2%

PRESS = 1871377678   R-Sq(pred) = 0,00%


Analysis of Variance

Source          DF         SS        MS     F      P
Regression       2  194805558  97402779  3,89  0,096
Residual Error   5  125248642  25049728
Total            7  320054200


Durbin-Watson statistic = 1,01415
```