

TESIS
PEMODELAN REGRESI SEMIPARAMETRIK BIRESPON
DENGAN ESTIMATOR *SPLINE* TERBOBOTI

MODELLING OF BIRESPONSE SEMIPARAMETRIC
*WITH WEIGHTED *SPLINE* ESTIMATOR*

AHMAD YANI

H062192001



PROGRAM STUDI MAGISTER STATISTIKA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
2023

TESIS
PEMODELAN REGRESI SEMIPARAMETRIK BIRESPON
DENGAN ESTIMATOR *SPLINE* TERBOBOTI

MODELLING OF BIRESPONSE SEMIPARAMETRIC
*WITH WEIGHTED *SPLINE* ESTIMATOR*

AHMAD YANI
H062192001



PROGRAM STUDI MAGISTER STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
2023

**PEMODELAN REGRESI SEMIPARAMETRIK BIRESPON
DENGAN ESTIMATOR *SPLINE* TERBOBOTI**

Tesis
sebagai salah satu syarat untuk mencapai gelar magister

Program Studi Magister Statistika

Disusun dan diajukan oleh

AHMAD YANI
H062192001

kepada

**PROGRAM STUDI MAGISTER STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
2023**

TESIS**PEMODELAN REGRESI SEMIPARAMETRIK BIRESPO
DENGAN ESTIMATOR SPLINE TERBOBOTI****AHMAD YANI
H062192001**

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka
Penyelesaian Program Studi Magister Statistika Fakultas
Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin
pada tanggal 1 Maret 2023
dan dinyatakan telah memenuhi syarat kelulusan

Menyetujui

Pembimbing Utama



Dr. Anna Islamiyati, S.Si., M. Si.
NIP. 19770808 200501 2 002

Pembimbing Pendamping



Dr. Nurtiti Sunusi, S.Si., M. Si.
NIP. 19720117 199703 2 002

Ketua Program Studi
Magister Statistika,

Dr. Dr. Georgina M. Tinungki, M.Si
NIP. 19620926 198702 2 001

Dekan Fakultas Matematika dan
Ilmu Pengetahuan Alam
Universitas Hasanuddin

Dr. Eng. Amiruddin, S.Si., M.Si.
NIP. 19720515 199702 1 002

PERNYATAAN KEASLIAN TESIS DAN PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa, tesis berjudul "Pemodelan Regresi Semiparametrik Birespon Dengan Estimator *Spline* Terboboti" adalah benar karya saya dengan arahan dari komisi pembimbing (Dr. Anna Islamiyati, S.Si., M.Si dan Dr. Nurtiti Sunusi, S.Si., M. Si.). Karya ilmiah ini belum diajukan dan tidak sedang diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka tesis ini. Sebagian dari isi tesis ini telah dipublikasikan di Jurnal (International Journal of Academic and Applied Research (IJAAR) ISSN: 2643-9603 Vol. 7 Issue 1, Januari - 2023, Pages: 150-153) sebagai artikel dengan judul "Estimation of Biresponse Semiparametric Regression Model with Weighted Spline Truncated".

Dengan ini saya melimpahkan hak cipta dari karya tulis saya berupa tesis ini kepada Universitas Hasanuddin.

Makassar, 1 Maret 2023



AHMAD YANI

NIM. H062192001

UCAPAN TERIMA KASIH

Puji syukur atas kehadiran Tuhan Yang Maha Esa atas limpahan rahmat dan karunia-Nya sehingga penulis dapat menyusun dan menyelesaikan tesis ini. Penulis menyadari sepenuhnya bahwa apa yang dikemukakan dalam tesis ini masih jauh dari kesempurnaan yang merupakan sebagai akibat dari keterbatasan kemampuan serta berbagai kesulitan yang penulis hadapi dalam penyusunan tesis ini.

Penulis memanjatkan doa kepada Tuhan Yang Maha Esa agar memberikan rahmat-Nya kepada pihak yang banyak membantu dalam penyelesaian tesis ini. Penulis juga percaya tesis ini dapat selesai bukan hanya dengan kekuatan pikiran penulis semata akan tetapi karena bantuan dari berbagai pihak juga, baik selama proses perkuliahan bahkan sampai proses pengerjaan tesis di Program Magister Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Hasanuddin. Namun demikian, penulis dengan senang hati menerima kritik dan saran yang bersifat membangun dari pembaca karya tulis ini demi sempurnanya tesis ini.

Terima kasih yang tak terhingga kepada orang tua dan saudara/saudariku tercinta atas doa yang tak pernah putus, dukungan serta segala kebaikan mereka yang sampai kapan pun takkan pernah bisa terbalaskan atas kasih sayang yang tiada henti dalam penyelesaian tesis ini. Selanjutnya, saya ingin menyampaikan juga rasa hormat dan terima kasih kepada:

1. **Dr. Eng. Amiruddin, M.Si.** selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam beserta seluruh jajarannya.
2. **Dr. Nurtiti Sunusi, M.Si.** selaku Ketua Departemen Statistika yang menjadi Pembimbing Pertama yang telah bersabar dan bersedia meluangkan waktunya untuk membimbing penulis dalam penyelesaian tesis ini. Terima kasih untuk dukungan, motivasi dan kemudahan yang diberikan kepada penulis..
3. **Dr. Dr. Georgina Maria Tinungki, M.Si.** selaku Ketua Program Studi Magister Statistika yang memberikan ilmu, dukungan, dan motivasi kepada penulis dalam berbagai hal selama menjalani Pendidikan di Departemen Statistika.
4. **Dr. Anna Islamiyati, S.Si., M.Si** selaku Pembimbing Utama yang telah bersabar dan bersedia meluangkan waktunya untuk membimbing penulis dalam penyelesaian tesis ini. Terima kasih untuk dukungan, motivasi dan kemudahan yang diberikan kepada penulis.
5. **Dr. Nirwan Ilyas, M.Si** selaku penguji penulis yang telah bersedia memberikan masukan-masukan dan arahan dalam penyusunan tesis.
6. **Dr. Erna Tri Herdiani, M.Si** selaku penguji penulis yang telah bersedia memberikan masukan-masukan dan arahan dalam penyusunan tesis.

7. **Sri Astuti Thamrin S.Si., M.Stat., Ph.D** selaku penguji penulis yang telah bersedia memberikan masukan-masukan dan arahan dalam penyusunan tesis.
8. Teman – teman Mahasiswa Program Magister Statistika angkatan Angkatan Pertama, terima kasih untuk dukungan dan semangat yang diberikan kepada penulis.
9. Teman-teman di Laboratorium Komputer Statistika yang telah membantu penyelesaian tesis penulis.
10. Teman-teman yang selalu mengingatkan dan menyemagati dalam penyelesaian tesis penulis.

Semoga Allah SWT memberikan pahala yang berlipat ganda atas segala kebaikan yang telah diberikan kepada penulis dan semoga penulisan tesis ini bermanfaat bagi perkembangan ilmu pengetahuan dan teknologi, khususnya dalam dunia statistika dan data sains.

Makassar, 1 Maret 2023

Ahmad Yani

ABSTRAK

AHMAD YANI. **Pemodelan Regresi Semiparametrik Birespon dengan Estimator *spline* Terboboti** (dibimbing oleh Dr. Anna Islamiyati, S.Si., M.Si. dan Dr. Nurtiti Sunusi, S.Si., M.Si.).

Model regresi umumnya dibagi menjadi tiga yaitu regresi parametrik, regresi nonparametrik dan regresi semi parametrik. Regresi semiparametrik merupakan kombinasi dari regresi parametrik dan regresi nonparametrik, karena terdapat hubungan antara variabel prediktor dan variabel respon yang polanya diketahui dan pola yang tidak diketahui. Tujuan dari penelitian ini adalah untuk mendapatkan model terbaik dari data biresponse menggunakan regresi semiparametrik. Indeks pembangunan manusia (IPM) dan nilai investasi di Sulawesi Selatan mejadi data yang akan diaplikasikan ke dalam pemodelan regresi semiparametrik birespon. Pendekatan regresi parametrik menggunakan regresi linier, sedangkan pendekatan regresi nonparametrik menggunakan *spline truncated* terboboti. Penentuan model regresi parametrik terbaik didasarkan pada koefisien determinasi terbesar, sedangkan penentuan model regresi nonparametrik didasarkan pada kriteria GCV untuk mendapatkan model terbaik.

Kata kunci: Birespon ; GCV; Investasi; IPM; Semiparametrik; Spline.

ABSTRACT

AHMAD YANI. ***Modelling of Biresponse Semiparametric with Weighted Spline Estimator***
(supervised by Dr. Anna Islamiyati, S.Si., M.Si. and Dr. Nurtiti Sunusi, S.Si., M.Si.).

Regression model are generally divided into three: parametric regression, nonparametric regression and semiparametric regression. Semiparametric regression is a combination of parametric regression and nonparametric regression. Because there is a relationship between predictor variables and response variables with known patterns and unknown patterns. The purpose of this study is to obtain the best model of biresponse data using semiparametric regression. The Human Development Index (HDI) and investment value in South Sulawesi are the data that will be applied to biresponse semiparametric regression modelling. The parametric regression approach uses linear regression, while the nonparametric regression approach uses weighted truncated spline. Determination of the best parametric regression model is based on the largest coefficient of determination, while determination of the nonparametric regression model is based on the GCV criterion to get the best model.

Keywords: Biresponse; GCV; HDI; Investment; Semiparametric; Spline.

DAFTAR ISI

PERNYATAAN PENGAJUAN	ii
HALAMAN PENGESAHAN	iii
PERNYATAAN KEASLIAN TESIS.....	iv
UCAPAN TERIMA KASIH	v
ABSTRAK.....	vii
ABSTRACT	viii
DAFTAR ISI.....	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Tujuan Penelitian	3
1.4. Manfaat Penelitian	3
1.5. Batasan Masalah.....	3
BAB II TINJAUAN PUSTAKA	5
2.1. Regresi Parametrik.....	5
2.2. Regresi Nonparametrik	6
2.3. Regresi Semiparametrik.....	7
2.4. Regresi <i>Spline</i>	8
2.5. Analisis Korelasi	9
2.6. Indeks Komposit Pembangunan Sosial-Ekonomi	10
BAB III METODE PENELITIAN	12
3.1. Sumber Data	12
3.2. Variabel Penelitian.....	12
3.3. Metode Analisis.....	12
BAB IV HASIL DAN PEMBAHASAN.....	14
4.1. Penaksir Model Regresi Semiparametrik Birespon <i>Spline Truncated</i>	14
4.2. Aplikasi Model Semiparametrik.....	21
4.2.1. Deskripsi Data IPM dan PNI	21
4.2.2. Pemilihan Titik Knot Pada Setiap Prediktor yang Berpola Nonparametrik Secara Parsial	25
4.2.3. Model Regresi Semiparametrik Birespon pada Data IPM dan PNI	28

BAB V PENUTUP	32
5.1. Kesimpulan	32
5.2. Saran.....	33
DAFTAR PUSTAKA.....	35
LAMPIRAN	38

DAFTAR TABEL

Tabel 2. 1 Interpretasi Koefisien Korelasi.....	10
Tabel 4. 1 Ringkasan	24
Tabel 4. 2 Nilai GCV Model Spline Linear untuk X_1 dan Y_1	25
Tabel 4. 3 Nilai GCV Model <i>Spline</i> Linear untuk X_1 dan Y_2	26
Tabel 4. 4 Nilai GCV Model <i>Spline</i> Linear untuk X_2 dan Y_2	26
Tabel 4. 5 Nilai GCV Model <i>Spline</i> Linear untuk X_3 dan Y_2	26
Tabel 4. 6 Nilai GCV Model <i>Spline</i> Linear untuk X_4 dan Y_2	27
Tabel 4. 7 Hasil estimasi parameter semiparametrik regresi birespon pada respon Y_1 (IPN)	28
Tabel 4. 8 Hasil estimasi parameter semiparametrik regresi birespon pada respon Y_2 (PNI)	28

DAFTAR GAMBAR

Gambar 4. 1 Kurva regresi linier antara Y_1 dengan X_1, X_2, X_3 dan X_4	22
Gambar 4. 2 Kurva regresi linier antara Y_2 dengan X_1, X_2, X_3 dan X_4	23
Gambar 4. 3 Kurva regresi <i>spline</i> antara Y_1 dengan X_1 serta Y_2 dengan X_1, X_2, X_3 , dan X_4 .	29

DAFTAR LAMPIRAN

Lampiran 1. Data Penelitian Prediktor dan Respon.....	38
--	----

BAB I

PENDAHULUAN

1.1. Latar Belakang

Pendekatan regresi bertujuan untuk melihat pengaruh variable prediktor terhadap variable respon. Terdapat tiga jenis regresi berdasarkan pola data, yaitu regresi parametrik, regresi nonparametrik dan regresi semiparametrik. Pola data antara prediktor dengan respon menjadi dasar penggunaan ketiga regresi tersebut. Pola data dapat diketahui dari plot data awal, informasi dari penelitian sebelumnya, kajian teori yang sudah ada, atau asumsi awal yang dibuat oleh peneliti dengan alasan-alasan tertentu. Regresi parametrik digunakan pada saat pola data mengikuti suatu pola parametrik dan sebaliknya regresi nonparametrik digunakan pada saat pola data tidak mengikuti suatu pola parametrik. Adapun regresi semiparametrik merupakan gabungan dari regresi parametrik dan nonparametrik. Hal ini berarti bahwa regresi semiparametrik digunakan pada saat terdapat pola parametrik dan pola nonparametrik sehingga estimasi regresi semiparametrik ekuivalen dengan mengestimasi parameter pada komponen parametrik dan nonparametrik (Budiantara, 2006).

Beberapa penelitian regresi semiparametrik pada satu respon yang telah dikembangkan oleh peneliti, diantaranya (Budiantara, 2005) mengembangkan semiparametrik unirespon dengan *spline smoothing*. (Hu, 2005) mengembangkan regresi ridge dalam pendekatan regresi semiparametrik. Sugiantari dan (Sugiantari & Budiantara, 2013) telah mengembangkan model regresi semiparametrik unirespon dengan menggunakan *spline smoothing*. (Chandra & Rohmaniah, 2019) menggunakan semiparametrik Cox. Estimator lain yang dapat digunakan dalam kasus birespon adalah estimator *spline truncated* yang memiliki interpretasi visual lebih mudah. *Spline truncated* mempertimbangkan titik knot yang menunjukkan titik terjadinya pola perubahan pada data. Hal tersebut yang menjadi keunggulan dari *spline truncated* karena mampu menjelaskan beberapa pola perubahan yang dapat terjadi pada satu model (Islamiyati, 2017). Penggunaan *spline truncated* banyak digunakan dalam beberapa data riil karena keunggulan tersebut, diantaranya (Anggreni et al., 2018) mengaplikasikan pada data tuberculosis, (Ramdhani et al., 2020) mengembangkan penggunaannya pada data diabetes.

Seiring dengan perkembangan data, permasalahan jumlah respon tidak hanya terbatas pada satu, akan tetapi telah berkembang pada jumlah respon dua atau lebih. (Chamidah & Rifada, 2016) mengembangkan analisis data birespon dengan estimator local linier, (Fernandes et al., 2015) dengan *smoothing spline* pada data longitudinal birespon. (Islamiyati, 2020) mengembangkan *penalized spline* pada data longitudinal birespon, (Hidayat et al., 2020) dengan estimator campuran antara kernel dan deret fourier. Jika kasus respon yang lebih dari satu dan diasumsikan saling berkorelasi, maka proses analisis data tidak dapat dimodelkan dengan pendekatan biasa seperti pada kasus unirespon. (Budiantara et al., 1997) telah mengembangkan *spline* parsial terboboti untuk data yang saling berkorelasi. (Davies & Meise, 2008) mengembangkan *smoothing spline* terboboti. (Chamidah et al., 2012) mengembangkan *polynomial local* terboboti. (Islamiyati et al., 2020) mengembangkan *penalized spline* terboboti dengan matriks varians kovariansi pada kasus birespon.

Salah satu data yang dapat dimodelkan dengan regresi semiparametrik birespon adalah data Indeks Pembangunan Manusia (IPM) dan perkembangan nilai investasi (PNI). Hal tersebut disebabkan karena IPM terdiri dari beberapa komponen yang saling berkorelasi. Komponen IPM tersebut diantaranya angka harapan hidup, harapan lama sekolah, rata-rata lama sekolah dan pendapatan per kapita. Selanjutnya, IPM tersebut dapat dipengaruhi oleh faktor harapan lama sekolah dan rata-rata lama sekolah yang diasumsikan berpola parametrik serta faktor angka harapan hidup dan pendapatan per kapita diasumsikan berpola nonparametrik sehingga dapat dimodelkan dengan regresi semiparametrik. Dalam penelitian yang dilakukan oleh (Pratomo & Sumargo, 2016) ukuran keberhasilan pembangunan suatu negara atau wilayah dapat dilihat dari pembangunan sosial-ekonomi daerah tersebut. IPM memberikan gambaran yang baik tentang dimensi sosial namun belum cukup menggambarkan dimensi ekonomi. Salah satu tolak ukur dimensi ekonomi dapat dilihat melalui investasi. Investasi sangat berpengaruh terhadap pertumbuhan ekonomi suatu wilayah baik itu investasi fisik, investasi finansial maupun investasi sumber daya manusia.

Berdasarkan uraian tersebut, dalam penelitian ini mengkaji tentang pemodelan regresi semiparametrik birespon dengan estimator *spline* terboboti. Model akan diaplikasikan pada data IPM dan PNI untuk mendapatkan model

optimal yang terkait dengan permasalahan IPM dan PNI di Propinsi Sulawesi Selatan berdasarkan data tahun 2014-2019.

1.2. Rumusan Masalah

Berdasarkan latar belakang, maka diperoleh rumusan masalah sebagai berikut:

1. Bagaimana mengestimasi model regresi semiparametrik birespon dengan estimator *spline* terboboti?
2. Bagaimana memodelkan data IPM dan PNI dengan model regresi semiparametrik birespon dengan estimator *spline* terboboti?

1.3. Tujuan Penelitian

Berdasarkan rumusan masalah, tujuan penelitian ini adalah sebagai berikut:

1. Mendapatkan estimasi model regresi semiparametrik birespon dengan estimator *spline* terboboti.
2. Mendapatkan model data IPM dan PNI dengan model regresi semiparametrik birespon dengan estimator *spline* terbobot.

1.4. Manfaat Penelitian

Berdasarkan rumusan masalah, manfaat penelitian ini adalah sebagai berikut:

1. Memberikan kontribusi keilmuan khususnya pengembangan ilmu di bidang statistika yang dapat diterapkan pada persoalan sosial ekonomi masyarakat.
2. Hasil dari penelitian ini dapat dijadikan bahan rujukan untuk mengambil kebijakan dalam hal perencanaan target IPM dan PNI kedepannya di wilayah Sulawesi Selatan.

1.5. Batasan Masalah

Dalam penelitian ini peneliti memberikan batasan-batasan masalah sebagai berikut:

1. Titik knot yang terpilih berdasarkan metode generalized cross validation (GCV).
2. Metode estimasi parameter yang digunakan adalah *Weighted Least Square* (WLS).

3. Data yang digunakan adalah data indeks pembangunan manusia (IPM) dan perkembangan nilai investasi (PNI) di Sulawesi Selatan pada tahun 2014 sampai dengan 2019.

BAB II

TINJAUAN PUSTAKA

2.1. Regresi Parametrik

Regresi parametrik merupakan metode analisis statistika yang digunakan untuk mencari pola hubungan antara variabel prediktor dan variabel respon, dengan asumsi bentuk pola hubungan diketahui. Untuk dapat menggunakan pendekatan regresi parametrik, diperlukan pengetahuan masa lalu tentang karakteristik data yang dapat diselidiki. Hubungan antara variabel prediktor dan variabel respon dalam model regresi dapat berpola linear maupun berpola nonlinear dalam parameter (Draper & Smith, 1998). Misalkan diberikan model regresi parametrik linear multivariabel atau yang umum dikenal sebagai regresi linear berganda sebagai berikut

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi} + \varepsilon_i, i = 1, 2, \dots, n \quad (2.1)$$

dengan

y_i : variabel respon pada pengamatan ke $-i$

$x_{1i}, x_{2i}, \dots, x_{pi}$: variabel prediktor pada pengamatan ke $-i$

$\beta_0, \beta_1, \dots, \beta_p$: koefisien regresi

ε_i : error acak pada pengamatan ke $-i$ yang berdistribusi $N(0, \sigma)$.

Jika persamaan regresi parametrik di atas dinyatakan dalam bentuk matrik, maka persamaan regresi parametrik tersebut menjadi

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2.2)$$

dengan

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{p1} \\ 1 & x_{12} & x_{22} & \cdots & x_{p2} \\ 1 & x_{13} & x_{23} & \cdots & x_{p3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{pn} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} \text{ dan } \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Parameter-parameter pada model di atas dapat diduga menggunakan metode statistika. Beberapa penduga yang sering digunakan dalam dalam penentuan nilai parameter-parameter tersebut adalah metode *Last Square* atau *Maximum*

Likelihood Estimator (Wahba, 1990). Persamaan untuk penduga parameter β dinyatakan sebagai berikut

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (2.3)$$

2.2. Regresi Nonparametrik

Pada kondisi tertentu, terdapat data yang memiliki persebaran yang acak sehingga sulit menentukan kurva untuk data tersebut. Dalam kasus ini, dapat diasumsikan bahwa bentuk kurva regresi tidak diketahui, salah satu saran yang ditawarkan dalam mengatasi hal tersebut dengan menggunakan pendekatan regresi nonparametrik.

(Eubank, 1999) memberikan persamaan untuk model regresi nonparametrik sebagai berikut

$$y_i = f(t_i) + \varepsilon_i, i = 1, 2, \dots, n \quad (2.4)$$

dengan

y_i : variabel respon pada pengamatan ke $-i$

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

ε_i : error acak pada pengamatan ke $-i$ yang berdistribusi $N(0, \sigma)$.

Persamaan (2.4) dapat dituliskan dalam bentuk matriks sebagai berikut

$$Y = f(T) + \varepsilon \quad (2.5)$$

dengan

$$y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}, f(T) = \begin{bmatrix} f(t_1) \\ f(t_2) \\ f(t_3) \\ \vdots \\ f(t_n) \end{bmatrix} \text{ dan } \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Fungsi regresi f dapat ditaksir menggunakan regresi nonparametrik. Pendekatan regresi nonparametrik memiliki fleksibilitas yang tinggi, sehingga diharapkan data dapat mencari sendiri bentuk taksiran kurva regresinya tanpa dipengaruhi oleh faktor subjektivitas peneliti (Eubank, 1999).

2.3. Regresi Semiparametrik

Regresi semiparametrik merupakan salah satu bagian dari statistika inferensi yang digunakan untuk memodelkan hubungan antara variabel respon dan prediktor yang sebagian dari variabel tersebut diketahui bentuk polanya dan sebagian lainnya tidak diketahui bentuk polanya. Regresi semiparametrik mulai berkembang pesat sejak Wahba pada tahun 1990 mempublikasikan model hubungan semiparametrik dari penggunaan listrik pada negara bagian di Amerika Serikat, yang berpola parametrik linear dengan pendapatan dan berpola nonparametrik dengan temperatur. Misalkan terdapat data berpasangan (x_i, y_i, t_i) dan hubungan antara x_i, y_i dan t_i diasumsikan mengikuti model regresi berikut (Ruppert et al., 2003).

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} + f(t_i) + \varepsilon_i, i = 1, 2, \dots, n \quad (2.6)$$

dengan

y_i : variabel respon pada pengamatan ke $-i$

$x_{1i}, x_{2i}, \dots, x_{pi}$: variabel prediktor pada pengamatan ke $-i$

$\beta_0, \beta_1, \dots, \beta_p$: koefisien regresi

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

ε_i : error acak pada pengamatan ke $-i$ yang berdistribusi $N(0, \sigma)$.

Persamaan (2.4) dapat dituliskan dalam bentuk matriks sebagai berikut

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{f}(\mathbf{T}) + \boldsymbol{\varepsilon} \quad (2.7)$$

dengan

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{21} & \dots & x_{p1} \\ 1 & x_{12} & x_{22} & \dots & x_{p2} \\ 1 & x_{13} & x_{23} & \dots & x_{p3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \dots & x_{pn} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}, \mathbf{f}(\mathbf{T}) = \begin{bmatrix} f(t_1) \\ f(t_2) \\ f(t_3) \\ \vdots \\ f(t_n) \end{bmatrix} \text{ dan } \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Matriks pada persamaan (2.7) yang komponen parametriknya mempunyai hubungan linear dengan variabel respon dan terdapat komponen nonparametrik juga disebut sebagai *Partial Linear Model* (Model Linear Parsial) (Hardle et al., 2000).

2.4. Regresi *Spline*

Regresi *spline* merupakan regresi non parametrik yang terdiri atas potongan-potongan pangkat dengan suatu orde tertentu dan tersambung pada titik-titik ikat. Nilai absis dari titik ikat biasa disebut knot. Knot diartikan sebagai suatu titik fokus dalam fungsi *spline*, sehingga kurva yang dibentuk tersegmentasi pada titik tersebut (Tripena, 2011). Satu nilai knot optimal nantinya dipilih dari berbagai knot yang telah dicoba. Knot optimal inilah yang digunakan sebagai parameter untuk menduga model regresi.

Fungsi *spline* berorde p dapat dinyatakan sebagai berikut (Biedermann et al., 2009).

$$f(x_i) = \beta_0 x_i^0 + \beta_1 x_i^1 + \beta_2 x_i^2 + \dots + \beta_p x_i^p + \sum_{k=1}^q \beta_{p+k} (x_i - K_k)^{+p}, \quad (2.8)$$

$$i = 1, 2, \dots, n \text{ dan } k = 1, 2, \dots, q$$

dengan

$x_1, x_2, \dots, x_i$: variabel prediktor pada pengamatan ke $-i$

$\beta_0, \beta_1, \dots, \beta_p$: koefisien regresi

q : banyaknya knot

K_k : titik knot ke- k

$$(x_i - K_k)^{+p} = \begin{cases} (x_i - K_k)^p & , x \geq K_k \\ 0 & , x < K_k \end{cases}, a < K_1 < K_2 < \dots < K_k < b \quad (2.9)$$

dengan a dan b masing-masing merupakan nilai terkecil dan terbesar dari data. Tanda “+” pada persamaan (2.9) menunjukkan bahwa hanya hasil $(x_i - K_k)^p \geq 0$ yang akan diambil sedangkan yang nilainya < 0 dianggap bernilai 0.

Penduga parameter β pada titik knot K dengan metode kuadrat terkecil dapat dituliskan sebagai

$$\hat{\beta}_k = (\mathbf{X}_k^T \mathbf{X}_k)^{-1} \mathbf{X}_k^T \mathbf{Y} \quad (2.10)$$

dengan $\mathbf{X}$ adalah matriks variabel prediktor dan $\mathbf{Y}$ adalah matriks variabel respon.

Jika dimiliki sejumlah n pengamatan, maka matriks fungsi dituliskan sebagai berikut

$$\mathbf{Y} = \mathbf{f}(x) = \mathbf{X}\boldsymbol{\beta} \quad (2.11)$$

dengan

$$\mathbf{f}(x) = \begin{bmatrix} f(x_1) \\ f(x_2) \\ f(x_3) \\ \vdots \\ f(x_n) \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_1^1 & x_1^2 & \dots & x_1^p (x_1 - K_1)^{+p} & \dots & (x_1 - K_k)^{+p} \\ 1 & x_2^1 & x_2^2 & \dots & x_2^p (x_2 - K_2)^{+p} & \dots & (x_2 - K_k)^{+p} \\ 1 & x_3^1 & x_3^2 & \dots & x_3^p (x_3 - K_3)^{+p} & \dots & (x_3 - K_k)^{+p} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & x_n^1 & x_n^2 & \dots & x_n^p (x_n - K_n)^{+p} & \dots & (x_n - K_k)^{+p} \end{bmatrix} \text{ dan}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \\ \beta_{(p+1)} \\ \beta_{(p+2)} \\ \vdots \\ \beta_{(p+q)} \end{bmatrix}$$

Pada regresi *spline* antar segmen dipisahkan oleh knot. Oleh karenanya, letak dan jumlah knot akan menentukan kebaikan regresi *spline* pada data (Eubank, 1999). Knot optimal dipilih berdasarkan nilai *Generalized Cross Validation (GCV)* paling kecil (Spiriti et al., 2018), yang dapat dihitung dengan persamaan berikut (Takezawa, 2006).

$$GCV(K_1, \dots, K_r) = \frac{MSE(K_1, \dots, K_r)}{\left(n^{-1} \text{tr}(\mathbf{I} - \mathbf{C}(K_1, \dots, K_r))\right)^2} \quad (2.12)$$

dengan

MSE diperoleh dari persamaan

$$MSE(K_1, \dots, K_r) = n^{-1} \mathbf{Y}' (\mathbf{I} - \mathbf{C}(K_1, \dots, K_r))' (\mathbf{I} - \mathbf{C}(K_1, \dots, K_r)) \mathbf{Y} \quad (2.13)$$

$\mathbf{C}(K_1, \dots, K_r)$ adalah matriks heat yang memuat nilai K parameter isolasi, r adalah banyaknya variabel komponen nonparametrik (Tripena, 2013).

2.5. Analisis Korelasi

Analisis korelasi merupakan analisis yang digunakan untuk mengetahui seberapa besar hubungan dua variabel atau lebih. Korelasi memiliki arahan positif dan negatif. Arahan positif memiliki arti bahwa hubungan antar dua variabel atau lebih yang bersifat searah. Sedangkan arahan negatif memiliki arti hubungan antar dua variabel atau lebih bersifat berlawanan arah.

Indeks korelasi atau koefisien korelasi merupakan angka yang dapat digunakan untuk mengukur kekuatan atau besaran (kuat, lemah atau tidak ada) korelasi pada variabel penelitian yang diuji. Nilai-nilai atau batasan korelasi dikelompokkan ke dalam beberapa interval (Supardi, 2017). Klasifikasi korelasi sesuai tingkatannya tersebut disajikan dalam Tabel 2.1 sebagai berikut.

Tabel 2. 1 Interpretasi Koefisien Korelasi

Nilai Koefisien Korelasi	Interpretasi
$KK = 0$	Tidak ada korelasi
$0,00 < KK \leq 0,20$	Korelasi sangat rendah /lemah sekali
$0,21 < KK \leq 0,40$	Korelasi rendah/lemah
$0,41 < KK \leq 0,70$	Korelasi yang cukup berarti
$0,71 < KK \leq 0,90$	Korelasi yang tinggi, kuat
$0,91 < KK \leq 0,99$	Korelasi sangat tinggi, sangat diandalkan
$KK = 1$	Korelasi Sempurna

Besaran satuan indeks koefisien korelasi berada pada interval 0 (nol) sampai $(\pm) 1$ (satu). Nilai terendah adalah 0 sedangkan nilai tertinggi 1. Jika dalam perhitungan nilai korelasi didapatkan hasil yang lebih besar dari 1 atau lebih kecil daripada -1 maka dapat disimpulkan dalam perhitungan terdapat kesalahan.

2.6. Indeks Komposit Pembangunan Sosial-Ekonomi

Indeks komposit merupakan indikator gabungan yang disusun dari indikator-indikator tunggal yang dikombinasikan sedemikian rupa menjadi sebuah indeks. Idealnya, indeks komposit mengukur konsep multidimensi, kompleks dan dapat mencakup banyak aspek yang tidak ditangkap oleh satu indikator/indeks tunggal. Indeks komposit pembangunan sosial-ekonomi digunakan untuk membandingkan kemajuan sosial ekonomi antara negara atau wilayah yang berbeda. Negara atau wilayah dengan indeks yang lebih tinggi cenderung memiliki tingkat kehidupan yang lebih baik dan stabil, sedangkan negara atau wilayah dengan indeks yang lebih rendah mungkin mengalami masalah seperti kemiskinan, ketidaksetaraan dan ketidakstabilan sosial. Beberapa peneliti telah mengajukan indikator dalam pengukuran indeks komposit pembanguana sosial-ekonomi diantaranya, (Morrison, 2005) membangun *the Physical Quality of Life Index (PQLI)* dan United Nation Development Program (UNDP) membangun *Human Development Index (HDI)*

atau Indeks Pembangunan Manusia (IPM) yang kini banyak digunakan oleh negara-negara di dunia dengan landasan yang dibangun oleh (Haq, 1995).

(UNDP, 1995) mendefinisikan pembangunan manusia sebagai suatu proses untuk memperluas pilihan-pilihan bagi penduduk dalam hal pendapatan, kesehatan, pendidikan, lingkungan fisik, dan sebagainya. Empat hal pokok yang perlu di perhatikan dalam pembangunan manusia adalah produktivitas, pemerataan, kesinambungan dan pemberdayaan. Indeks Pembangunan Manusia (IPM) merupakan indikator komposit yang mengukur tiga dimensi pokok pembangunan manusia dan dinilai mampu mencerminkan kemampuan dasar (*basic capabilities*) penduduk. Ketiga kemampuan dasar itu adalah panjang umur dan sehat yang dapat diukur menggunakan angka harapan hidup waktu lahir (AHH), berpengetahuan dan berketerampilan yang dapat diukur menggunakan harapan lama sekolah (HLS) dan rata-rata lama sekolah (RLS), serta akses terhadap sumber daya yang dibutuhkan untuk mencapai standar hidup layak yang dapat diukur menggunakan pengeluaran perkapita.

Indikator yang dapat digunakan untuk melihat pembangunan ekonomi adalah investasi. Untuk menghasilkan output, selain membutuhkan bahan baku dan tenaga kerja juga diperlukan barang modal yang digunakan dalam proses produksi. Dengan investasi diharapkan dapat menambah stok kapital dan pada gilirannya akan meningkatkan kapasitas produksi. Dengan demikian, output dan pendapatan masyarakat akan meningkat. Dalam jangka panjang akumulasi investasi dapat mendorong perkembangan berbagai aktivitas ekonomi sehingga meningkatkan pertumbuhan ekonomi di suatu wilayah/negara (BPS, 2019).